

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО"**

Кваліфікаційна наукова
праця на правах рукопису

СКУЛИШ МАРІЯ АНАТОЛІВНА

УДК 621.391

ДИСЕРТАЦІЯ

**Моделі та методи керування обслуговуванням гібридних сервісів в
телекомунікаційному середовищі з використанням хмарних ресурсів**

05.12.02 – телекомунікаційні системи та мережі

(шифр і назва спеціальності)

Подається на здобуття наукового ступеня
доктора технічних наук

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

(підпис, ініціали та прізвище здобувача)

Науковий консультант
Глоба Лариса Сергіївна
д.т.н., професор

Київ – 2019

АНОТАЦІЯ

Скулиш М. А. Моделі та методи керування обслуговуванням гібридних сервісів у телекомунікаційному середовищі з використанням хмарних ресурсів. - Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.12.02 «Телекомунікаційні системи та мережі» (172 – Телекомунікації та радіотехніка). – Національний технічний університет України "Київський політехнічний інститут ім. Ігоря Сікорського" Міністерства освіти і науки України, Київ, 2019.

Дисертаційну роботу присвячено проблематиці, пов'язаній із організацією взаємодії елементів телекомунікаційної системи (ТКС) та віртуального зовнішнього обчислювального середовища з метою підвищення якості обслуговування гібридних телекомунікаційних сервісів за рахунок адаптивного вибору обчислювальних ресурсів, систематизації та удосконалення методів керування потоками між обслуговуючими сутностями та процесом обслуговування у гетерогенному телекомунікаційному середовищі (ГТС) для гарантування якісного обслуговування кінцевих користувачів та контрольованого використання обчислювальних ресурсів в умовах ускладнення структури сервісів, підвищення вимог до показників якості обслуговування та неперервного зростанням об'ємів трафіку.

Динаміка розвитку телекомунікаційних та інформаційних технологій обумовила збільшення різноманітності послуг, які надаються операторами зв'язку, що створює передумови для побудови вискоєфективної гетерогенної інформаційно-телекомунікаційної інфраструктури для забезпечення гарантованих показників якості обслуговування сервісів. Разом з тим, підтримка функціонування такої надскладної інфраструктури вимагає значного обсягу обчислювальних ресурсів для керування процесами обслуговування сервісів. У таких умовах світовою тенденцією вирішення проблематики якісного надання широкого спектру послуг будь де та в будь-

який час стає створення технологій п'ятого покоління, які широко використовують можливості «хмарних» віддалених дата центрів, що в свою чергу потребує розробки нових принципів, моделей та методів керування процесом обслуговування гібридних сервісів у гетерогенному телекомунікаційному середовищі. Напрямок розвитку інформаційно-телекомунікаційної інфраструктури, впровадження новітніх інформаційних технологій, зокрема, грид- та хмарних технологій, також став пріоритетним для побудови інформаційного суспільства в Україні.

В роботі продемонстровано передумови для виникнення протиріччя, оскільки досі немає єдиної концепції, моделей та методів організації контролю та наскрізного керування процесом обслуговування гібридних сервісів, які б враховували особливості розгортання мережі дата центрів, обмеження мережевих інтерфейсів, які поєднують територіально-розподілені дата центри, особливості логічної структури віртуальної обчислювальної мережі, яка розгортається поверх фізичної інфраструктури, а також систем оркестрування і організації взаємодії всіх елементів та підсистем ГТС, що забезпечують процес надання гібридних телекомунікаційних послуг. Це призводить до неефективного використання обчислювальних ресурсів, які забезпечують гетерогенне середовище обслуговування сучасних сервісів.

На сьогодні дане протиріччя можливо розв'язати шляхом розробки нових принципів, математичних моделей та методів обслуговування потоків у ГТС, резервування ресурсів у віртуалізованому середовищі для обслуговування надмірного потоку навантаження, прогнозування необхідного віртуального обчислювального ресурсу у хмарі на основі довгострокової та короткострокової статистики, що забезпечуватиме гнучкість та масштабованість підсистеми керування процесом обслуговування гібридних телекомунікаційних сервісів.

У дисертаційній роботі вирішено важливу науково-технічну проблему, пов'язану з розробкою наукових основ організації взаємодії елементів ТКС та віртуального зовнішнього обчислювального середовища, які базуються на

систематизації та розвитку методів керування гібридними телекомунікаційними сервісами, процесом їх обслуговування у ГТС для уникнення зниження якості обслуговування під час сплесків перевантаження, утримання показники якості на заданому рівні, зменшення кількості відмов за рахунок скорочення часу реагування підсистеми керування телекомунікаційною та хмарною частинами гетерогенного середовища.

Перший розділ роботи містить огляд літературних джерел за темою дисертації, проведення аналізу сучасного стану та тенденцій розвитку, особливостей функціонування ГТС, визначено фактори впливу на показники якості обслуговування застосованих у його складі технологій віртуалізації мережевих функцій, визначено та систематизовано підходи щодо забезпечення якості обслуговування кінцевих користувачів. Досліджено особливості мобільних мереж зв'язку, а також тенденції розвитку систем зв'язку, структурні зміни у сервісах, що надаються кінцевим користувачам, еволюційні зміни у підсистемах забезпечення функціонування ТКС, зокрема зростання впливу показників якості функціонування обчислювальних підсистем на роботу ТКС оператору зв'язку.

В другому розділі було досліджено моделі контролю якості обслуговування гібридних телекомунікаційних сервісів керування та показників функціонування системи, які впливають на якість обслуговування кінцевого користувача. Виявлено особливості процедур обслуговування користувачів у мобільній мережі, а також у підсистемах ГТС, проаналізовано вплив віртуалізації мережевих функцій на процедури забезпечення якості обслуговування кінцевих користувачів гібридних телекомунікаційних сервісів.

Третій розділ присвячено дослідженню та удосконаленню підходів щодо побудови архітектури ГТС, організації процесу контролю якості обслуговування гібридних телекомунікаційних сервісів керування, визначенню показників ефективності використання ресурсів у підсистемах ГТС, впливу віртуалізації мережевих функцій на процедури забезпечення

якості обслуговування кінцевих користувачів гібридними телекомунікаційними сервісами. Представлено удосконалений принцип керування телекомунікаційною системою, який на відміну від SDN-мереж, дозволяє розмістити у хмарному середовищі як всі функції керування службовим трафіком, так і процеси взаємодії між ними у відповідності до діючих стандартів, передаючи тільки результуючий керівний вплив на фізичні елементи, які здійснюють передавання інформаційних потоків, що зменшує обсяги службового трафіку, який циркулює телекомунікаційною мережею.

У четвертому розділі роботи запропоновано принцип резервування обчислювальних ресурсів у віртуалізованому гетерогенному середовищі для обслуговування надмірного потоку навантаження на елементи ГТС, який дозволить забезпечити незмінними показники якості обслуговування потоків гібридних сервісів застосовуючи динамічно сформовану схему використання обчислювальних ресурсів. Для цього розроблено математичну модель задачі визначення максимально допустимого обсягу навантаження з гарантуванням рівня QoS для вузла обслуговування у ГТС, яка враховує ергодичний розподіл кількості заявок, граничну затримку, мінімізує втрати запитів в системі через нестачу обчислювальних ресурсів обслуговування та дозволяє розрахувати верхню межу допустимого навантаження під час планування інтенсивності вхідного потоку для обслуговування у обчислювальному вузлі. Для вибору потужності обслуговуючого пристрою розроблено математичну модель, яка гарантує утримання процесу обслуговування на заданому рівні якості та вирізняється тим, що замість балансування навантаження між сталою кількістю обслуговуючих пристроїв враховує масштабованість віртуальних обчислювальних ресурсів вузлів обслуговування, мінімізуючи кількість цих пристроїв та втрати часу на балансування між обслуговуючими сутностями.

У п'ятому розділі проведено дослідження щодо визначення обсягу необхідної кількості обчислювальних ресурсів під час надання сервісів

кінцевому користувачу в ГТС. Запропоновано принцип керування необхідним обсягом віртуального обчислювального ресурсу у хмарі на основі статистичних даних (довгострокових та короткострокових), який використовує дані системи моніторингу для розрахунку конфігурацій обчислювальних ресурсів та розкладу їх використання, періодично контролює достатність обраної конфігурації, забезпечуючи таким чином показники якості, гнучкості та масштабованості підсистеми керування процесом обслуговування гібридних телекомунікаційних сервісів, а також ефективне використання віртуальних та фізичних обчислювальних ресурсів. Розроблено метод побудови розкладу залучення обчислювальних ресурсів для обслуговування гібридних телекомунікаційних сервісів з заданим рівнем QoS, сутність якого полягає у визначенні тренду зміни обсягів навантаження протягом доби на основі довгострокової статистики за рахунок ітераційної оцінки його статистичних характеристик на інтервалах часу різної довжини, що дозволяє із урахуванням очікуваного навантаження сформувати розклад зміни конфігурації обчислювальних ресурсів вузла обслуговування на добу, зменшуючи таким чином час на реконфігурацію ТКС. Вдосконалено авторегресійний метод із ковзковим математичним очікуванням за рахунок визначення на основі довгострокової статистики інтервалу аналізу поточних статистичних даних, що забезпечує достовірність прогнозу із заданою ймовірністю та мінімізує кількість точок аналізу, а відповідно й операцій під час перевірки достатності обчислювальних ресурсів у режимі реального часу для обслуговування поточного навантаження в ТКС.

В шостому розділі досліджено запропоновані модифікації компонентів підсистеми PCRF (Policy and Charging Rules Function), нові процедури організації їх взаємодії між собою та з підсистемами керування віртуалізованим середовищем для забезпечення контролю якості обслуговування гібридних телекомунікаційних сервісів, проведено імітаційне та математичне моделювання процесів керування сервісами у ГТС, що дозволило підтвердити можливість запобігти погіршенню показників QoS на

рівні абонентського пристрою кінцевого користувача та зменшити кількість відмов за рахунок скорочення часу реагування підсистеми керування телекомунікаційною та хмарною частинами гетерогенного середовища. Досліджено запропоновані архітектурні рішення щодо побудови телекомунікаційної мережі з віртуалізацією мережевих функцій та забезпеченням заданого рівня QoS, керуванням обчислювальними ресурсами ТКС, що показало доцільність проведення перерозподілу навантаження між групою фізичних та віртуальних вузлів обслуговування, покращення коефіцієнту використання обчислювальних ресурсів, а саме зменшення відсотку запитів, які було втрачено через перевищення допустимого часу обслуговування, а також зменшення відсотку часу низької завантаженості обчислювальних ресурсів. Проведено математичне моделювання нових процедур та алгоритмів, зокрема алгоритму планування обсягу ресурсів ГТС, процедур організації взаємодії підсистем ТКС, керування віртуалізованим телекомунікаційним середовищем із забезпеченням заданого рівня QoS та плануванням обсягу необхідних обчислювальних ресурсів для ефективної роботи ГТС. Результати моделювання запропонованих технічних рішень показали збільшення ймовірності утримання ресурсів обслуговування у заданих межах. Досліджено ефективність застосування запропонованих принципів, математичних моделей та методів керування процесом обслуговування гібридних телекомунікаційних сервісів в діючій мережі оператора зв'язку.

Ключові слова: віртуалізація мережевих функцій, ресурси віртуальних мережевих функцій, гібридний телекомунікаційний сервіс, гетерогенне телекомунікаційне середовище, QoS, програмно-керовані мережі.

ABSTRACT

Skulysh M.A. Models and methods of managing the service of hybrid services in a telecommunication environment using cloud resources. - Proficiency scientific treatise on the rights of the manuscript.

A thesis submitted in fulfilment of the Doctor of Engineering Science degree in technical sciences on specialty 05.12.02 «Telecommunication systems and networks» (172 - Telecommunications and Radioengineering). – National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" of Ministry for Education and Science of Ukraine, Kyiv, 2019.

The dissertation is devoted to the problem, connects with the organization of the interaction of telecommunication system (TCS) elements and the virtual external computing environment. The main goal is to improve the quality of service of hybrid telecommunication services due to the adaptive choice of computing resources, systematization and improvement of flow management methods between serving entities and the service process in a heterogeneous telecommunication environment (GTE). It will ensure high-quality service of end-users and controlled use of computing resources in conditions of service structures complication, increased requirements for service quality indicators and continuous increase of traffic volumes.

The dynamics of telecommunication and information technologies development has led to an increase in the diversity of services provided by telecommunication operators. This creates the prerequisites for the construction of a highly effective heterogeneous information and telecommunication infrastructure to ensure guaranteed service quality indicators. However, supporting the operation of such a complex infrastructure requires a large amount of computing resources to manage service processes. In such conditions, in order to solve the problems of providing a wide range of services anywhere and at any time, the tendency of creating the fifth-generation technologies widely use the opportunities of "cloud" remote data centers becomes worldwide relevant. This requires the development of new principles, models and methods of managing the service of hybrid services in

a heterogeneous telecommunication environment. The direction of the development of information and telecommunication infrastructure, introduction of the latest information technologies, in particular, grid and cloud technologies, has also become a priority for the construction of an information society in Ukraine.

The preconditions for the emergence of contradiction are demonstrated. There is still no single concept, models and methods for organizing control and cross-channel management of the service of hybrid services that take into account the features of data center network deployment, the limitation of network interfaces, which combine geographically distributed data centers, features of the logical structure of virtual computer network, which unfolds over the physical infrastructure, as well as systems of orchestration and organization the interaction of all GTE elements and subsystems providing the process of hybrid communication services. This leads to inefficient use of computing resources that provide a heterogeneous environment for servicing modern services.

Today, this contradiction can be solved by: development of new principles, mathematical models and methods of flow service in the GTE; resource reservation in a virtualized environment for servicing excessive load flow; prediction of the required virtual cloud computing resource on the basis of long-term and short-term statistics. This will ensure the flexibility and scalability of the hybrid telecommunication services process control subsystem.

The dissertation presents the solution of the actual scientific problem related to the development of the scientific basis for the organization of TCS and virtual external computing environment elements interaction. These elements are based on the systematization and development of managing hybrid telecommunication services methods, the process of their maintenance in the GTE to avoid reducing the quality of service during load jumps, the maintenance of quality indicators on the given level, the reduction of failures number by reducing the response time of the telecommunication and cloud parts of the heterogenetic environment control subsystem.

The first section of the paper contains an overview of literary sources on the dissertation topic, the analysis of the current state and development trends, the features of the GTE functioning. The factors influencing the quality of servicing of the network functionalities virtualization technologies applied in its structure, the approaches to quality assurance of end users are defined and systematized. The features of mobile communication networks and trends of communication systems development, structural changes in services provided to end-users, evolutionary changes in the subsystems of ensuring the functioning of TCS, in particular, the increase of the influence of the computing subsystems quality performance indicators on the work of TCS providers are investigated.

In the second section, models of quality control of hybrid telecommunication management services and system performance indicators that affect the quality of service of the end-user are investigated. The features of user servicing procedures in the mobile network, as well as in the subsystems of the GTS are revealed, the influence of virtualization of network functions on the procedures for ensuring the quality of service of end users of hybrid telecommunication services is analyzed.

The third section is devoted to the research and development of approaches to building the architecture of the GTE, the organization of the process of service quality control of hybrid telecommunication management services. The indicators of the resources usage efficiency in the GTE subsystems, the impact of the network functions virtualization on the procedures for ensuring the quality of end users serving by hybrid telecommunication services are determined. The advanced principle of telecommunication system management is presented. Unlike SDN networks, it allows placing all the functions of management of service traffic and interaction processes between them in the cloud environment in accordance with the existing standards. It only transfers the resulting management influence to the physical elements that transmit information flows, which reduces the amount of service traffic circulates over the telecommunication network.

The fourth section of the paper proposes the principle of computing resource reservation in a virtualized heterogeneous environment to maintain an excessive

load flow on the GTE elements, which will ensure that the quality of service of the hybrid services flows is unchanged by applying a dynamically generated computational resource scheme. For this purpose, a mathematical model of the problem of determining the maximum permissible load volume with guaranteeing the level of QoS for the service node in the GTE was developed. It takes into account the ergodic distribution of the applications number and the limiting delay, minimizes the loss of queries in the system due to the lack of computing service resources and allows calculating the upper limit of the allowable load when planning the intensity of the input stream in the computing node. To select the power of the service unit, a mathematical model that guarantees maintenance of the serving process at a given level of quality was developed. It is distinguished by the fact that, instead of balancing the load between a steady number of service units, the scalability of virtual computing resources of service nodes is taken into account, minimizing the number of these devices and loss of time for balancing between service providers.

In the fifth section, a study was conducted to determine the amount of required computing resources when providing services to the end user in the GTE. The principle of management of the required amount of virtual computing resource in the cloud based on statistical data (long-term and short-term), which uses monitoring system data to calculate the computational resources configurations and their use scheduling is proposed. It periodically monitors the adequacy of the selected configuration, thus providing quality indicators, flexibility and scalability of the subsystem of hybrid telecommunications services managing process, as well as the efficient use of virtual and physical computing resources. A method of constructing a schedule of computational resources involvement for the maintenance of hybrid telecommunication services with a given level of QoS is developed. Its essence is to determine the trend of changes in load volumes during the day based on long-term statistics due to the iterative evaluation of its statistical characteristics at different lengths time intervals. This allows formulating a schedule for changing the configuration of the computing resources of the service

node per day, taking into account the expected load, thus reducing the time for reconfiguring the TCS. The auto-regression method with rolling mathematical expectation by determining interval of analysis of the current statistics based on long-term statistics. It ensures the reliability of the forecast with the given probability and minimizes the number of analysis points, respectively operations when checking the adequacy of computing resources in real-time to maintain the current load in the TCS.

The sixth section examines the proposed modifications of the PCRf (Policy and Charging Rules Function) components, the new procedures for organizing their interaction with each other and with the virtualization environment control subsystems to ensure the quality control of the hybrid telecommunication services. The simulation and mathematical modeling of service management processes in the GTE was conducted. It allowed to confirm the possibility of preventing QoS indicators deterioration at the level of the end user subscriber device and to reduce the number of failures by reducing the response time of the subsystem of telecommunication and cloud control parts of the heterogeneous environment. The proposed architectural decisions concerning the construction of a telecommunication network with the virtualization of network functions, the provision of a given QoS level and management of computing TCS resources were investigated. The research has shown the feasibility of reallocating the load between a group of physical and virtual service nodes, improving the use of computing resources coefficient, namely, reducing the percentage of requests that were lost due to exceeding the allowable maintenance time, as well as reducing the percentage of low computing time load. The mathematical modeling of new procedures and algorithms, in particular, the algorithm of planning the volume of GTE resources, procedures for organizing the interaction of TCS subsystems, management of a virtualized telecommunication environment with the provision of a given QoS level and planning of the volume of necessary computing resources for the efficient operation of the GTE was carried out. The results of simulation of the proposed technical solutions showed an increase in the probability of

maintenance of service resources within the specified limits. The efficiency of application of the proposed principles, mathematical models and methods of managing the process of serving hybrid telecommunication services in the operator's network was investigated.

Key words: network function virtualization, resources of virtual network functions, hybrid telecommunication service, heterogeneous telecommunication environment, QoS, software-defined networks.

Список публікацій здобувача:

Наукові праці, в яких опубліковані основні наукові результати дисертації:

Монографія:

1. Л.С. Глоба, О.М. Дяденко, А.Ю. Пилипенко, М.А. Скулиш. Математичні методи аналізу та керування телекомунікаційними мережами. К.: Інститут обдарованої дитини НАПН України. 2017. 234с.

Статті у виданнях іноземних держав або у фахових виданнях України, які включені до міжнародних наукометричних баз

2. Globa L., Slukysh M. Nodal routing with traffic classification. *Polish association for knowlage management Series: Studies&Proceedings*. Bydgoszcz, Polska. 2011. №42. P. 37-46. (Польща). (Наукове фахове видання з технічних наук).

3. Globa L., Slukysh M. Calculation functions of SDN controller for wireless backhaul infrastructure. *Acta Electrotechnica et Informatica*. 2017. Volume 17. Number 4. pp. 14-18. (Словачія). (Наукове фахове видання з технічних наук).

4. Скулиш М. А., Суліма С. В. Гібридна система управління ресурсами для віртуалізованих мережевих функцій. *Радіоелектроніка, інформатика, управління*, p-ISSN 1607-3274, Запоріжжя, 2017, № 1, С.16-24. (Наукове фахове видання України з технічних наук. Входить до міжнародних

наукометричних баз: DOI, DOAJ, British Library, eLibrary.ru / РИНЦ, Google Scholar, Index Copernicus та ін.)

5. Скулиш М. А., Суліма С.. В. Система управління ресурсами в центрах обробки даних оператора мережі мобільного зв'язку. *Visnyk NTUU KPI Serii – Radiotekhnika Radioaparobuduvannia*, ISSN: 2310-0397, e-ISSN: 2310-0389. 2017. Iss. 68. С. 27-32. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: DOAJ, Google Scholar, WEB of Science, Index Copernicus, eLibrary.ru / РИНЦ та ін.)

6. Skulysh. M. Application of the "Endless train" method for the SDN controller OpenDayLight. *Visnyk NTUU KPI Serii – Radiotekhnika Radioaparobuduvannia*, ISSN: 2310-0397, e-ISSN: 2310-0389. 2017. Iss. 69. P. 32-38. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: DOAJ, Google Scholar, WEB of Science, Index Copernicus, eLibrary.ru / РИНЦ та ін.)

7. Скулиш. М.А. Заставенко А.А. Метод розподілу ресурсів сервера оператора мобільного зв'язку. *Visnyk NTUU KPI Serii – Radiotekhnika Radioaparobuduvannia*, ISSN: 2310-0397, e-ISSN: 2310-0389. 2015. № 60. С. 35-45. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: DOAJ, Google Scholar, WEB of Science, Index Copernicus, eLibrary.ru / РИНЦ та ін.)

8. Skulysh M., Globa L. Planning the loading of data centers' resources based on download statistics. *Visnyk NTUU KPI Serii – Radiotekhnika Radioaparobuduvannia*, ISSN: 2310-0397, e-ISSN: 2310-0389. 2016. Iss 65. P. 62-72. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: DOAJ, Google Scholar, WEB of Science, Index Copernicus, eLibrary.ru / РИНЦ та ін.)

Статті у наукових фахових виданнях:

9. Скулиш М. А. Організація роботи групи серверів для забезпечення потреб розподіленої системи тарифікації послуг. *Наукові записки Українського науково-дослідного інституту зв'язку*. 2014. №5(33). С. 56-64.
10. Скулиш М. А. Метод складання розкладу залучення ресурсів для високонавантажених інформаційних систем. *Наукові записки Українського науково-дослідного інституту зв'язку*. 2014. №6(34). С.65-70.
11. Скулиш. М.А. Метод згладжування вхідного навантаження на сервер on-line тарифікації. *Вісник Національного технічного університету «Львівська політехніка» Радіoeлектроніка та телекомунікації*. Львів. 2014. №796. С. 97-103
12. Скулиш. М.А. Методи контролю перевантаженнями в процесі багатоетапної обробки заявок. *Наукові записки Українського науково-дослідного інституту зв'язку*. 2015. №1(35). С.81-86.
13. Скулиш М. А., Суліма С. В. Організація управління в багатоетапних системах масового обслуговування. *Вісник Національного технічного університету «Львівська політехніка» Радіoeлектроніка та телекомунікації*. Львів. 2015. № 818. С. 80–85.
14. Skulysh M., Klimovych O. Method of LTE functional units organization with evolved packet core virtualization. *Telecommunication Sciences*. 2015. No. 4. P. 38–45.
15. Skulysh M., Globa L., Sulima S. Model for Efficient Allocation of Network Functions in Hybrid Environment. *Telecommunication Sciences*. 2016. Volume 7. Number 1, P. 39-45.
16. Скулиш М.А., Заставенко А.А. Задача розподілу абонентського навантаження між базовими станціями з підтримкою SDR. *Телекомунікаційні та інформаційні технології*. 2016. №4, С. 42-49.
17. Скулиш. М.А. Заставенко А.А. Метод контролю якості обробки інформаційних потоків у мережі 5G. *Вісник Національного технічного*

університету «Львівська політехніка» *Радіoeлектроніка та телекомунікації*. Львів. 2016. №849. С.265-273.

18. Глоба Л.С., Вольвач Є. О., Скулиш М.А. Система керування якістю обслуговування у складних гетерогенних телекомунікаційних системах. *Системи управління, навігації та зв'язку. Збірник наукових праць*. Полтава: ПНТУ, 2017. Т. 6 (46). С. 158-162.

19. Скулиш М.А. Формування зони обслуговування прийомо-передавальної станції в залежності від змінного вхідного навантаження. *Вісник Національного університету «Львівська політехніка». Радіoeлектроніка та телекомунікації*. Львів. 2017. № 885. С. 50-54.

20. Skulysh M., Sulima S. Hybrid resource provisioning system for telecommunication network. *Сучасні інформаційні системи*. Харків. 2018. Том 2, № 1. С. 47-51.

21. Скулиш М.А., Романов О.І., Глоба Л.С. Принцип обслуговування потоків у гетерогенному телекомунікаційному середовищі. *Вчені записки таврійського національного університету імені В.І. Вернадського, Серія: Технічні науки*. 2018. Том 29 (68) № 2. С. 92-98.

22. Тимченко І.О., Скулиш М.А., Глоба Л.С. Концепції побудови сучасних мереж. *Радиоэлектроника и информатика = Radioelectronics & informatics : науч. журн. Харьк. нац. ун-т радиоэлектроники*. Харків: ХНТУРЭ. 2018. №1. С. 25-29.

23. Скулиш М.А., Романов О.І., Глоба Л.С. Принцип резервування ресурсів у віртуалізованому середовищі для контролю показників обслуговування. *Вісник Інженерної академії України*. 2018. С. 197-202

24. Шилов Ф., Скулиш М., Сафарян А. Дослідження ефективності методу оптимального вибору обчислювальних ресурсів для білінгових систем. *Системи управління, навігації та зв'язку*. 2018. Вип. 3. С. 147-152.

25. Скулиш М.А. Математична модель пошуку оптимального обсягу ресурсів віртуального телекомунікаційного вузла обслуговування. *Сучасні інформаційні системи*. Харків. 2018. Том 2, № 2. С. 30-34.

26. Скулиш М.А. Романов О.І., Нестеренко М.М. Принцип прогнозування необхідного віртуального ресурсу хмарної системи для оператора мобільного зв'язку. *Збірник наукових праць ВІПІ*. Київ. 2018. Вип. №2. С. 113 – 119.

Патенти України на корисну модель:

27. Скулиш М.А. Глоба Л.С. Ватула А.С, Марчук А.П. Система обробки мультисервісних інформаційних потоків. *Патент на корисну модель №85488*. Патент опубліковано 25.11.2013, бюл. № 22/2013.

28. Скулиш М.А. Глоба Л.С. Ватула А.С, Марчук А.П. Спосіб зваженого кругового обслуговування черг мультисервісних інформаційних потоків. *Патент на корисну модель №85487*. Патент опубліковано 25.11.2013, бюл. № 22/2013.

Матеріали та тези наукових конференцій, які індексуються у Scopus:

29. Globa L. S., Yermakova K. O., Kushnyr V. V., Skulysh M. A. Paralle computing usage for computation of the queues QoS parameters. *20th International Crimean Conference "Microwave & Telecommunication Technology"*. 2010. P. 466-467 DOI: 10.1109/CRMICO.2010.5632345

30. Globa L. S., Skulysh M. A. Efficiency increase for the choice of the virtual connection direction. *20th International Crimean Conference "Microwave & Telecommunication Technology"*. 2010. С. 502 - 503 DOI: 10.1109/CRMICO.2010.5632474

31. Globa L. S., Skulysh M. A. Improving efficiency service of queue in nodal centers. *21st International Crimean Conference "Microwave & Telecommunication Technology"*. 2011. P. 461-462.

32. Globa L. S., Skulysh M. A. Estimation scheme for improvement of weighted round robin algorithm for requests service in the queues. *22nd*

International Crimean Conference "Microwave & Telecommunication Technology". 2012. P.393 – 394.

33. Globa L. S., Skulysh M. A. Method of the analysis of power consumption dependence from service of a number of processes. *23rd International Crimean Conference "Microwave & Telecommunication Technology". 2013. C.432 – 433.*

34. Globa L. S., Skulysh M. A., Reverchuk A.V. Method of increase of fault tolerance of mobile communication systems at the moments of peak load. *23rd International Crimean Conference "Microwave & Telecommunication Technology". 2013. C.533 – 534.*

35. M.A. Skulysh. Method of increase of efficiency of assistance to the victims according to the integrated number 112. *23rd International Crimean Conference "Microwave & Telecommunication Technology". 2013. C. 545 – 546.*

36. Globa L. S., Skulysh M. A., Reverchuk A.V. Control strategy of the input stream on the online charging system in peak load moments. *24th International Crimean Conference Microwave & Telecommunication Technology. 2014. P. 312 – 313.*

37. Globa L. S., Skulysh M. A., Pidgurska T., Reverchuk A. Managing of incoming stream applications in online charging system. *Telecommunications (BIHTEL), 2014 X International Symposium on. IEEE. 2014. P. 1-6.*

38. Skulysh M., Sulima S. Management of Multiple Stage Queuing Systems. *CADSM 2015 : 13-th International conference, 24–27 February 2015 : conference proceedings. Lviv–Polyana. 2015. P. 431– 434.*

39. Globa L. S., Skulysh M. A., Zastavenko A. The method of resources allocation for processing requests in online charging system. *Experience of Designing and Application of CAD Systems in Microelectronics (CADSM) 2015 : 13-th International conference, 24–27 February 2015 : conference proceedings. Lviv–Polyana. 2015. P. 211-213.*

40. Skulysh M., Klimovych O. Approach to virtualization of Evolved Packet Core Network Functions. *Experience of Designing and Application of CAD*

Systems in Microelectronics (CADSM), 2015 13th International Conference The. – IEEE, 2015. – pp. 193-195.

41. Skulysh M., Sulima S. Service deployment aspects in the systems with network function virtualization. *Radio Electronics & Info Communications (UkrMiCo)*, 2016 International Conference : conference proceedings. IEEE, 2016. – pp. 1-7.

42. Globa L. S., Skulysh M. A., Sulima S. Method for resource allocation of virtualized network functions in hybrid environment. *2016 IEEE International Black Sea Conference on Communications and Networking* : conference proceedings. 2016. P: 1 - 5, DOI: 10.1109/BlackSeaCom.2016.7901546.

43. Skulysh M. The Method of Computing Organization in High Loaded SDN Controller System. *Experience of Designing and Application of CAD Systems in Microelectronics (CADSM) 2017* : 13-th International conference, 24–27 February 2017 : conference proceedings. Lviv–Polyana, 2017. P. 431– 434.

44. Skulysh M. The method of resources involvement scheduling based on the long-term statistics ensuring quality and performance parameters. *Radio Electronics & Info Communications (UkrMiCo) 2017 International Conference* : conference proceedings. 2017. P. 1-4.

45. Skulysh M., Sulima S., Sunduchkov K., Savchuk B. Sampling for direct search method of all system implementations developed according to the requirements of the technical design specification. *Problems of Infocommunications. Science and Technology (PIC S&T) 4th International Scientific-Practical Conference* : conference proceedings. 2017 P: 83 – 86.

46. Skulysh M., Romanov O. The structure of a mobile provider network with network functions virtualization. *14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)* : conference proceedings. 2018. P: 1032 – 1034.

Матеріали та тези міжнародних наукових конференцій:

47. Скулиш М. А. Суліма С. В. Організація управління в багатоетапних системах масового обслуговування. *Проблеми телекомунікацій ПТ-2015* : зб. матеріалів доп. учасн. IX Міжнар. наук.-техн. конф. Київ. 2015. С. 319–321.
48. Скулиш М. А., Суліма С. В. Управління ресурсами в віртуалізованих мережах оператора мобільного зв'язку. *Інфокомунікації – сучасність та майбутнє* : V Міжнар. наук.- практ. Конф., 29-30 жовтня 2015 : матеріали конференції. Одеса, 2015. С.84–88.
49. Скулиш М. А., Суліма С. В. Алгоритм відображення та планування віртуалізованих функцій в мережі мобільного зв'язку. *Проблеми телекомунікацій ПТ-2016* : 10-а міжнародна науково-технічна конференція, 19–22 квітня 2016 : матеріали конференції. — Київ, 2016. — сс. 372–374.
50. Skulysh M., Sulima S. Service deployment aspects in the systems with network function virtualization. *RadioElectronics&InfoCommunications* : First International Conference, 11–16 September 2016 : conference proceedings. — Kyiv, 2016. — P. 1–7.
51. Скулиш М.А., Заставенко А.А., Яндович І.Є. Оцінка залежності використання ресурсів від вхідного навантаження на сервері мобільного оператора. *Проблеми телекомунікацій ПТ-2014* : 8-а міжнародна науково-технічна конференція, матеріали конференції. Київ, 2014 . С. 202–206.
52. Глоба Л.С., Скулиш М.А., Реверчук А.В. Схема згладжування вхідного навантаження на сервер мобільного оператора. *Проблеми телекомунікацій ПТ-2014* : 8-а міжнародна науково-технічна конференція : матеріали конференції. — Київ, 2014 . — сс. 46-49.
53. Тимченко І.О., Скулиш М.А. Метод відновлення вузла у мережах NFV. *Проблеми телекомунікацій ПТ-2017* : 11-а міжнародна науково-технічна конференція, 18–21 квітня 2017 : матеріали конференції. Київ. 2017 . С. 262–264.
54. Запорожец Д.Б., Скулиш М.А, Анализ методов прогнозирования с использованием нейронной сети. *Проблеми телекомунікацій ПТ-2018* : 10-а

міжнародна науково-технічна конференція : матеріали конференції. Київ. 2018 . С. 266-268.

55. Скулиш М.А., Чиж В.В. Впровадження віртуалізації мережевих функцій на мережі платформи передачі SMS – повідомлень. *Проблеми телекомунікацій ПТ-2018* : 10-а міжнародна науково-технічна конференція : матеріали конференції. Київ, 2018 . С. 272-275.

56. Белокур Б.В., Скулиш М.А. Планування ресурсів гетерогенного телекомунікаційного середовища на базі аналізу статистичних даних. *Проблеми телекомунікацій ПТ-2018* : 10-а міжнародна науково-технічна конференція : матеріали конференції. Київ, 2018 . С. 279-281.

57. Скулиш М.А., Запорожец Д.Б. Сравнительный анализ методов краткосрочного прогнозирования сетевого трафика. *Проблеми телекомунікацій ПТ-2018* : 10-а міжнародна науково-технічна конференція : матеріали конференції. Київ, 2018 . С. 310-312.

58. Скулиш М.А., Нездвезький В.С. Аналіз засобів моделювання для мережевої СМО із динамічно змінюваними параметрами вузлів обслуговування. *Проблеми телекомунікацій ПТ-2018* : 10-а міжнародна науково-технічна конференція, матеріали конференції. Київ, 2018. С. 310-312.

59. Скулиш М.А. Проблеми впровадження системи екстреної допомоги населенню за єдиним номером 112. *Проблеми телекомунікацій ПТ-2013* : 7-а міжнародна науково-технічна конференція : матеріали конференції. Київ. 2013 . С. 138-140.

60. Скулиш М.А., Глоба Л.С., Система комплексного керування інформаційними потоками в комутаційному центрі. *Проблеми телекомунікацій ПТ-2013* : 7-а міжнародна науково-технічна конференція : матеріали конференції. Київ. 2013. С. 141-144.

61. Гайдаенко О.В., Скулиш М.А. Существующие основные подходы по методам оценки качества передачи речевой информации в сетях IP. *Електроніка-2013* : IV Міжнародна конференція молодих вчених : матеріали конференції. Київ. 2013. С. 302-305.

62. Марчук А.П., Скулиш М.А. Математична модель методу оптимізації обслуговування викликів у «Системі 112». *Електроніка-2013 : IV Міжнародна конференція молодих вчених : матеріали конференції*. Київ. 2013. С. 302-305.

63. Скулиш М.А., Бабій О.С. Проблеми побудови наукових інформаційних порталів. *Современные проблемы радиотехники и телекоммуникаций РТ-2011 : VII Международная молодежная научно-техническая конференция : матеріали конференції*. г. Севастополь. 2011. С. 174-175.

64. Скулиш М.А., Марчук А.П. Метод оптимізації обслуговування викликів у «Системах-112». *Современные проблемы радиотехники и телекоммуникаций РТ-2011 : IX Международная молодежная научно-техническая конференция : матеріали конференції*. Севастополь. 2013 г. С. 98-99.

Статті у інших виданнях.

65. Глоба Л.С., Скулиш М.А. Керування трафіком в мультисервісному комутаційному центрі. *Telecommunication Sciences*. 2011. Т. 2. №1. С. 30-38.

66. Скулиш М.А., Суліма С.В. Метод управління ресурсами в датацентрах оператора мережі мобільного зв'язку. *Radioelectronics & Informatics Journal* (ISSN 1563-0064). Харків. 2015. № 3. С. 8-13.

67. Скулиш М.А. Заставенко А.А. Пилипенко А.Ю. Метод вибору контейнера для міграції віртуальної машини в хмарне сховище даних. *Вісник Харківського національного університету імені В.Н. Каразіна, серія "Математичне моделювання. Інформаційні технології. Автоматизовані системи управління"*. 2016 р. Випуск 31. С.47-58.

68. Скулиш М. А. Метод організації функцій контролеру SDN «Нескінчений потяг». *Radioelectronics & Informatics Journal* (ISSN 1563-0064). Харків. 2016. № 3. С. 21–28.

ЗМІСТ

ВСТУП	33
РОЗДІЛ 1	45
Аналіз сучасного стану проблем керування інформаційно-телекомунікаційними системами	45
1.1. Мережі NGN як основа для 5G. Особливості системи керування.....	48
1.2. Система організації хмарних обчислень. Аналіз стандартів для забезпечення керування телекомунікаційними системами	57
1.2.1. Особливості керування сервісами із застосуванням хмарних обчислень.....	59
1.2.2. Загальні вимоги до керування сервісами у гетерогенному інформаційно-телекомунікаційному середовищі.....	64
1.2.3. Функціональні вимоги до керування сервісами в хмарній системі керування телекомунікаціями	67
1.3. Мережа LTE формальний опис структури , основні підсистеми та інтерфейси.....	73
1.3.1. Протоколи доступу LTE E-UTRAN	73
1.3.2. Інтерфейси між вузловими елементами в мережах стандарту LTE	79
1.4. Особливості віртуалізації підсистем мобільного зв'язку LTE	81
1.4.1. Гетерогенне телекомунікаційне середовище	81
1.4.2. Віртуалізація маршрутизатора на границі локальної мережі оператора мобільно зв'язку CPE (vE-CPE) - функціональний опис	83
1.4.3. Віртуалізація порогового маршрутизатора оператора (vPE) - функціональний опис	84
1.4.4. Граф переадресації VNF.....	87
1.4.5. Віртуалізація базової мобільної мережі та IMS.....	92
1.4.6. Віртуальна мережева функція як сервіс	100
1.5. Архітектура мережі LTE з віртуалізацією мережевих функцій у гетерогенній хмарній інфраструктурі	104
Висновки	110

РОЗДІЛ 2	112
Моделі Контролю якості обслуговування кінцевих користувачів	112
2.1. Моделі контролю якості обслуговування.....	112
2.1.1. Вимоги до якості в 5G мережі	115
2.2. Контроль якості у системі з віртуалізованою системою керування.....	118
2.2.1. Забезпечення якості в мережах LTE без віртуалізації	118
2.2.2. Вимоги до NFV і SDN інтеграції в мобільних мережах	120
2.2.3. Показники якості послуг передачі даних	122
2.2.4. Аналіз методів забезпечення параметрів QoS в сервісно-орієнтованій архітектурі LTE	126
Основні функції мережі LTE, що належать до керування якістю	128
2.2.5. Реалізація концепції QoS у мережах LTE.....	130
2.2.6. Процедури виділеного каналу для протоколу доступу E-UTRAN з S5/S8 на основі PMIP	132
2.3. Вплив віртуалізації мережевих функцій на процедури забезпечення якості обслуговування.	135
2.4. Модель керування енергозбереженням телекомунікаційної системи, навантаженням та ресурсами обслуговування.....	138
Висновки	145
РОЗДІЛ 3	146
Система керування гетерогенною телекомунікаційною системою	146
3.1. Цілі і завдання керування на телекомунікаційних мережах.....	146
3.2. Системи керування на телекомунікаційних мережах	148
3.2.1. Узагальнена модель процесу керування телекомунікаційною мережею	149
3.2.2. Класифікація систем керування телекомунікаційними мережами.....	152
3.2.3. Функції організаційних систем керування	154
3.3. Програмно визначені мережі	157
3.3.1. Передумови впровадження SDN	159

3.3.2. SDN контролери	161
3.3.3. Інтеграція SDN та NFV.....	165
3.4. Віртуалізація в мережах мобільного зв'язку	167
3.4.1. Інтеграція SDN в мобільних мережах.....	168
3.5. Метод організації ресурсів віртуалізованих мережеских функцій EPC	170
3.5.1. Підхід до віртуалізації EPC.....	170
3.5.2. Організація функціональних вузлів мережі LTE.....	177
3.5.3. Передача повідомлень управління при віртуалізації EPC	181
Висновки	184
РОЗДІЛ 4	186
Резервування ресурсів у віртуалізованому середовищі для обслуговування надмірного потоку навантаження на елементи ГТС	186
4.1. Процедура гарантування заданої якості обслуговування	186
4.2. Задача формування вхідного потоку навантаження.....	190
4.3. Задача вибору потужності вузла обслуговування	193
Задача вибору потужності вузла обслуговування набуває особливої.....	193
4.4. Задачі розподілу ресурсів при обслуговуванні потоку гібридних телекомунікаційних сервісів на сервері онлайн тарифікації послуг запитів.....	196
4.4.1. Структура системи тарифікації. Сучасні системи обробки та тарифікації викликів в мобільних мережах зв'язку	197
4.4.2. Процес обробки викликів в мобільних мережах зв'язку	204
4.4.3. Обробка викликів на сервері тарифікації	208
4.4.4. Задача розподілу ресурсів сервера.....	211
4.4.5. Оцінка ефективності методу розподілу ресурсів	214
4.5. Методи розподілу технічних ресурсів для гібридного телекомунікаційного потоку	217
4.5.1. Задача визначення розподілу кількості заявок які у поточний момент часу знаходяться на обслуговуванні в системі.	219

4.5.2. Передумови створення системи керування вхідним потоком на вузол хмарного середовища.....	222
4.5.3. Постановка задачі методу керування вхідним потоком на сервер мобільного зв'язку.	230
4.5.4. Моделювання методу динамічного керування вхідним навантаженням.	236
4.6. Задача реконфігурації гетерогенного телекомунікаційного середовища ..	238
4.6.1. Метод відображення віртуальних вузлів на фізичні вузли	238
4.6.2. Алгоритм динамічного розподілу мережевих ресурсів	243
4.6.3. Принцип роботи методу відображення і планування мережевих функцій.....	245
4.7. Метод “нескінченний потяг” організації обчислень у хмарному середовищі для гібридних телекомунікаційних сервісів	248
Висновки	253
РОЗДІЛ 5	255
прогнозування необхідного віртуального ресурсу у хмарі для забезпечення ефективної роботи ГТС	255
5.1. Система динамічного керування ресурсами мережі	257
5.2. Розрахунок конфігурацій системи обслуговування гібридних телекомунікаційних сервісів.	258
5.2.1. Вибір найкращої конфігурації, що забезпечує енергоефективну роботу ГТС	261
5.3. Метод складання розкладу використання конфігурацій	264
5.4. Оцінка ефективності статистичного методу розподілу кількості хмарних ресурсів.....	268
5.5. Контроль достатності зарезервованих ресурсів.....	269
5.5.1. Алгоритм пошуку інтервалу контролю достатності ресурсів.....	272
5.6. Модуль моніторингу та вимірювання величин.....	274

5.7. Розробка розкладів включення обладнання. Метод визначення моменту включення додаткового технічного засобу	280
5.7.1. Механізми моніторингу платформи	285
5.8. Методологія забезпечення ефективного обслуговування ібридних телекомунікаційних сервісів у ГТС.....	287
Висновки	291
РОЗДІЛ 6	292
ІМІТАЦІЙНІ МОДЕЛІ ЗАПРОПОНОВАНИХ МЕТОДІВ	292
6.1. Дослідження ефективності методу оптимального вибору обчислювальних ресурсів для білінгових систем.....	292
6.1.1. Метод формування вхідного потоку навантаження для ефективного використання ресурсів обслуговування.....	293
6.2. Імітаційна модель MATLAB методу формування вхідного потоку навантаження для ефективного використання ресурсів обслуговування	305
6.3. Імітаційна модель системи онлайн тарифікації з додатковим сервером в середовищі GPSS.....	309
6.4. Моделювання задачі розподілу ресурсів між гібридними телекомунікаційними сервісами	318
6.5. Імітаційна модель системи онлайн тарифікації із змінним ресурсом обслуговування.....	331
6.6. Імітаційна модель для методу вибору потужності обслуговуючого пристрою	337
6.7. Імітаційна модель MATLAB для методу побудови розкладу залучення ресурсів на основі довгострокової статистики із забезпеченням показників якості та енергоефективності.....	344
6.8. Імітаційна модель MATLAB для методу короткострокового прогнозування навантаження.....	353
Висновки	358

ВИСНОВКИ..... 360

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... 365

ДОДАТОК 1.....397

ДОДАТОК 2.....411

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

3G – Third Generation

3GPP – 3rd Generation Partnership Project

AIPN – All IP Network

ANI – Application Network Interface

ARPU

AS – Access stratum

ATM – Asynchronous Transfer Mode

AuC – Authentication Center

BBU – Baseband unit

BSS – Business Support System

CDMA – Code Division Multiple Access

CN – Core Network

CPE – Customer Premises Equipment

CSP – Cloud Service Provider

DL – Down Link

DPI – Deep Packet Inspection

DSL – Digital Subscriber Line

E2E – End-to-End

eNB – Evolved Node B

EPS – Evolved Packet Core

EPS – Evolved Packet System

eTOM – Enhanced Telecom Operations Map

ETSI – European Telecommunications Standards Institute

FCAPS – Fault, Configuration, Accounting, Performance, and Security

FE – Functional Element

FI – Functional Interface

FW – Firewall

GPRS – General Packet Radio Service

HLR – Home Location Register

HSS – Home Subscriber Server

IEEE – Institute of Electrical and Electronics Engineers

IP – Internet Protocol

IPS – Intrusion Prevention System

ISDN – Integrated Services Digital Network

ITU – International Telecommunication Union

KQI – Key Quality Indicator

LTE – Long-Term Evolution

MAC – Medium Access Control layer

MIMO – Multiple Input Multiple Output

MPLS – Multiprotocol Label Switching

NACFs – Network Attachment Control Functions

NAPT – Network Address Port Translation

NAS - non-Access Stratum

NFV – Network Functions Virtualization

NFVI – NFV Infrastructure

NG-FW – Next Generation Firewall

NGN – Next Generation Networks

OFDM – Orthogonal Frequency-Division Multiplexing

OSS – Operation Support System

PDCP – Packet Data Convergence Protocol

PDN – Public data network

PE – Provider Edge

PGW – Packet Data Network Gateway

PHY – Physical layer

PNF – Physical Network Function

QoS – Quality of Service

RAB - Radio Access Bearer

RACF – Resource and Admission Control Function

RAN – Radio Access Network

RAT – Remote Administration Tool

RLC – Radio Link Control

RRC – Radio Resource Control

RRM - radio resource management

SAE – System Architecture Evolution

SCF – Service Control Functions

SDN – Software-defined networks

SGSN – Serving GPRS Support Node

SGW – Serving Gateway

SIP – Session Initiation Protocol

SLA – Service Level Agreement

SMI – Service Management Interface

UE – User equipment

UNI – User-network interface

VE-CPE – virtualized enterprise CPE

VeCPU – Virtual Central Processing Unit

VM – Virtual Machine

VNF – Virtual Network Function

VNFaaS – Virtual Network Functions as a Service

VPLS — Virtual Private LAN Service

WOC – WAN Optimization Controller

БС – базова станція

IT – Information Technologies

MME – Mobility Management Entity

MCE – Міжнародний союз електрозв'язку

РУО – Рівень керування обслуговуванням

РУР – Рівень керування ресурсами

ТфЗК – Телефонна мережа загального користування

ГТС – гетерогенне телекомунікаційне середовище

ВСТУП

Актуальність теми. Телекомунікаційна мережа оператора мобільного зв'язку сьогодні – це організована система, яка включає в себе спеціальне обладнання, яке обслуговується, спостерігається та керується з операційних дата центрів, де встановлено обчислювальні сервери та відповідне програмне забезпечення, що обслуговує чисельні інформаційні та службові потоки. Сучасні технології SDN, NFV, SDR, CloudRAN та інші стрімко розвиваються, повномасштабне їх впровадження призведе до повної залежності працездатності телекомунікаційної мережі від роботи інформаційно-обчислювального середовища.

Постачальники послуг тепер віртуалізують частини своєї мережі, що суттєво впливає на рішення для аналізу та оцінки параметрів функціонування, які використовуються для забезпечення роботи мережі. Однією з областей розвитку систем забезпечення якості обслуговування абонентів в умовах часткової віртуалізації систем зв'язку є контроль показників якості на всіх етапах надання сервісів абонентам при використанні як телекомунікаційного обладнання так і обчислювальних систем.

Спостерігається інтеграція телекомунікаційних систем (ТКС) та розподіленого обчислювального середовища, в результаті утворюється єдине гетерогенне середовище обслуговування телекомунікаційних сервісів, в якому є можливість контролювати процес обслуговування інформаційних потоків на кожному етапі та забезпечити відповідність високим стандартам якості. У той же час досі немає єдиної концепції, моделей та методів організації контролю та наскрізного керування, які б враховували особливості розгортання мережі дата центрів, обмеження мережевих інтерфейсів, які їх поєднують, особливості логічної структури віртуальної обчислювальної мережі, яка розгортається поверх фізичної інфраструктури, а також систем оркестрування і організації взаємодії всіх елементів та

підсистем, які забезпечують процес надання телекомунікаційних послуг. Це призводить до неефективного використання ресурсів, що утворюють гетерогенне середовище обслуговування телекомунікаційних сервісів.

Особливості розвитку інформаційно-телекомунікаційних мереж:

- У зв'язку зі зближенням індустрії інформаційних та комунікаційних технологій, телекомунікаційна інфраструктура все більше використовує хмарні обчислення. Телекомунікаційні оператори надають хмарні сервіси, а також застосовують технології хмарних обчислень для оптимізації своїх телекомунікаційних платформ та систем підтримки (ITU-T M.3371).

- Спостерігається відсутність зворотного зв'язку між якістю послуг, які надаються, та організацією процесів взаємодії у гетерогенному телекомунікаційному середовищі, як наслідок хаотичне завантаження обчислювальних та телекомунікаційних ресурсів, які забезпечують розподілену гетерогенну систему обслуговування.

- Потреба у гнучких моделях та методах керування якістю обслуговування гібридних телекомунікаційних сервісів, які б використовували переваги гетерогенного телекомунікаційного середовища та враховували особливості обчислювальних процесів.

Впровадження технологій програмно керованих мереж потребує впровадження нових моделей теорії масового обслуговування для оцінки параметрів функціонування системи, своєчасного виконання обчислювальних операцій для забезпечення потреб ТКС.

Через відсутність методологічної бази для організації роботи гетерогенного телекомунікаційного середовища (ГТС), його ресурси використовуються хаотично, задачі оптимізації вирішені частково або локально, що призводить до погіршення контролю та забезпечення показників якості послуг для кінцевих користувачів. Розроблені в роботі моделі та методи є складовими єдиної архітектури контролю та керування ресурсами і потоками на рівні провайдера мобільного зв'язку.

Таким чином, створення і наукове обґрунтування комплексної методології керування процесом обслуговування у гетерогенному телекомунікаційному середовищі з метою підвищення якості процесу обслуговування гібридних телекомунікаційних сервісів є актуальною науково-технічною проблемою.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась згідно із планами науково-дослідних робіт кафедри інформаційно-телекомунікаційних мереж КПІ ім. Ігоря Сікорського у рамках держбюджетних тем №2333-п «Мультиагентна система інтеграції інформаційних ресурсів та обробки інформації в розподіленому інформаційно-телекомунікаційному середовищі» (Номер держреєстрації 0110U002415,), №2418-п «Розробка міждисциплінарного комплексу розподілених обчислень на базі веб-сервісів» (Номер державної реєстрації – 0111U000332), №2619-п «Гетерогенне середовище з динамічною архітектурою для високопродуктивної обробки інформації в розподілених інформаційних системах» (Номер державної реєстрації – 0113U001627), №2861-п «Технологія обробки сервісів з інтеграцією інформаційних ресурсів в системах підтримки операційної діяльності підприємств зв'язку» (Номер державної реєстрації – 0115U000217), №2117-п «Технологія побудови динамічних реєстрів електронних інформаційних ресурсів та засобів їх ефективної обробки у датацентрах гетерогенної структури» (Номер державної реєстрації – 0118U003522) та госпдоговірної роботи: «Розробка системної інфраструктури єдиного інформаційного середовища даних антарктичних досліджень. Оптимізація системи супутникового зв'язку для забезпечення ефективного функціонування антарктичної станції «Академік Вернадський» (Договір №Н/Н-2013-5-08.07.2013).

Мета і задачі дослідження. Дисертаційна робота присвячена вирішенню важливої науково-технічної проблеми, пов'язаної з розробкою наукових основ організації взаємодії елементів ТКС та обчислювального середовища, які базуються на систематизації методів керування

телекомунікаційними сервісами та процесом їх обслуговування у гетерогенному телекомунікаційному середовищі для забезпечення якісного обслуговування кінцевих користувачів та контрольованого використання обчислювальних ресурсів обслуговування.

Метою дослідження є підвищення якості обслуговування гібридних телекомунікаційних сервісів у гетерогенному телекомунікаційному середовищі за рахунок організації потоків між обслуговуючими сутностями та адаптивного вибору ресурсів обслуговування сервісів.

Для досягнення мети дослідження було поставлено та вирішено такі основні задачі:

1. Проаналізувати особливості функціонування ГТС, визначити фактори впливу на показники якості обслуговування застосованих у його складі технологій віртуалізації мережевих функцій, визначити та систематизувати підходи щодо забезпечення якості обслуговування кінцевих користувачів.

2. Розробити методологію забезпечення якості обслуговування, яка дозволяє керувати процесом обслуговування гібридних телекомунікаційних сервісів у гетерогенному телекомунікаційному середовищі.

3. Розробити принцип обслуговування потоків у ГТС, який забезпечить гнучкість та масштабованість системи керування процесом обслуговування гібридних телекомунікаційних сервісів у ГТС.

4. Запропонувати принцип резервування ресурсів у віртуалізованому середовищі для обслуговування надмірного потоку навантаження на елементи ГТС, який дозволить утримувати незмінними показники якості обслуговування потоків гібридних сервісів при заданій схемі використання ресурсів.

5. Розробити математичну модель пошуку оптимального навантаження на систему обслуговування із раннім виявленням перевантаження, яка дозволить забезпечити обслуговування із заданими показниками QoS, а саме ймовірністю успішного обслуговування, часом обслуговування гібридних

телекомунікаційних сервісів та утриманням коефіцієнту використання ресурсів у заданих межах.

6. Розробити математичну модель пошуку оптимального обсягу ресурсу додаткового віртуального вузла обслуговування, який забезпечить якісне обслуговування навантаження із заданими показниками якості обслуговування гібридних телекомунікаційних сервісів.

7. Запропонувати принцип прогнозування обсягу необхідного віртуального ресурсу у хмарі для забезпечення роботи ГТС, який дозволить, використовуючи дані системи моніторингу, розрахувати конфігурації системи обслуговування та розклад їх використання з метою забезпечення показників якості.

8. Розробити метод побудови розкладу вибору конфігурацій на основі довгострокової статистики, який враховує статистичні характеристики потоку гібридних телекомунікаційних сервісів та можливі конфігурації системи обслуговування.

9. Вдосконалити авторегресійний метод із ковзковим математичним очікуванням прогнозування навантаження для контролю достатності ресурсів обслуговування гібридних телекомунікаційних сервісів у ГТС в режимі реального часу.

10. Оцінити роботу запропонованих рішень.

Об'єктом дослідження є процес взаємодії гібридних телекомунікаційних сервісів під час обслуговування кінцевих користувачів у ГТС.

Предметом дослідження є моделі, методи, методики та інструменти керування процесом обслуговування гібридних телекомунікаційних сервісів ГТС.

Методи дослідження. Основні методи дослідження загальної проблеми – методи теорії систем та мереж масового обслуговування, теорії ймовірності, методи математичної статистики, теорії оптимізації, математичне моделювання фізико-теоретичні основи передачі даних і

радіозв'язку, теорії комп'ютерних систем і інформаційно-телекомунікаційних мереж, теорії алгоритмів тощо. За допомогою методів теорії масового обслуговування та теорії імовірності досліджується вплив дисциплін обробки викликів на роботу системи обробки телекомунікаційних сервісів оператора у гетерогенному хмарному середовищі. За допомогою математичного моделювання проводилася оцінка якості керування обробкою сервісів у комплексі програмно-керованих сутностей телекомунікаційної мережі. За допомогою теорії динамічного програмування та засобів теорії дослідження операцій було розв'язано ряд оптимізаційних задач пошуку найкращої конфігурації мережі.

Наукова новизна отриманих результатів. Наукова новизна одержаних результатів полягає у наступному:

1. Запропоновано методологію керування процесом обслуговування гібридних телекомунікаційних сервісів, яка на відміну від відомих використовує оригінальні принципи, математичні моделі та методи, що у сукупності дозволяють організувати одночасне гнучке керування як потоками сервісів, так і необхідними для цього обчислювальними ресурсами обслуговування (фізичними та хмарними), утворюючи гетерогенне телекомунікаційне середовище, яке гарантує задані показники якості обслуговування.

2. Вдосконалено принцип керування телекомунікаційною системою, який на відміну від принципу SDN-мереж, розміщує у хмарному середовищі як всі функції керування службовим трафіком, так і процеси взаємодії між ними у відповідності до діючих стандартів; передаючи тільки результуючий керівний вплив на фізичні елементи, які здійснюють передавання інформаційних потоків, що зменшує обсяги службового трафіку, який циркулює телекомунікаційною мережею.

3. Вперше запропоновано математичну модель задачі визначення максимально допустимого обсягу навантаження з гарантуванням рівня QoS для вузла обслуговування у гетерогенному телекомунікаційному середовищі,

яка враховує ергодичний розподіл кількості заявок, граничну затримку, мінімізує втрати запитів в системі через нестачу обчислювальних ресурсів обслуговування та дозволяє розрахувати верхню межу допустимого навантаження під час планування інтенсивності вхідного потоку для обслуговування на вузлі.

4. Вперше запропоновано математичну модель задачі вибору потужності обслуговуючого пристрою, яка гарантує утримання процесу обслуговування на заданому рівні якості, на відміну від існуючих підходів щодо балансування навантаження (коли для забезпечення зростаючого навантаження додаються додаткові обслуговуючі пристрої між якими виконується балансування запитів) враховує масштабованість віртуальних обчислювальних ресурсів вузлів обслуговування, що мінімізує кількість обслуговуючих пристроїв та втрати часу на балансування між обслуговуючими сутностями.

5. Запропоновано вперше принцип керування необхідним обсягом віртуального ресурсу у хмарі на основі статистичних даних (довгострокових та короткострокових), який використовує дані системи моніторингу для розрахунку конфігурацій обчислювальних ресурсів обслуговування та розкладу їх використання, періодично контролює достатність обраної конфігурації, забезпечуючи таким чином показники якості, гнучкість та масштабованість системи керування процесом обслуговування гібридних телекомунікаційних сервісів, а також ефективне використання віртуальних та фізичних обчислювальних ресурсів.

6. Запропоновано вперше метод побудови розкладу залучення обчислювальних ресурсів для обслуговування гібридних телекомунікаційних сервісів з заданим рівнем QoS, сутність якого полягає у визначенні тренду зміни обсягів навантаження протягом доби на основі довгострокової статистики за рахунок ітераційної оцінки його статистичних характеристик на інтервалах часу різної довжини, що дозволяє із урахуванням очікуваного навантаження сформувати розклад зміни конфігурації обчислювальних

ресурсів вузла обслуговування на добу, зменшуючи таким чином час на реконфігурацію телекомунікаційної системи.

7. Вдосконалено авторегресійний метод із ковзковим математичним очікуванням за рахунок визначення на основі довгострокової статистики інтервалу аналізу поточних статистичних даних, що забезпечує достовірність прогнозу із заданою ймовірністю та мінімізує кількість точок аналізу, а відповідно й операцій під час перевірки достатності обчислювальних ресурсів у режимі реального часу для обслуговування поточного навантаження в телекомунікаційній системі.

Практичне значення одержаних результатів.

1. Усі теоретичні розробки дисертаційної роботи автором доведено до конкретних архітектурних рішень, протоколів взаємодії та алгоритмів технології керування гібридними сервісами у телекомунікаційних системах нового покоління, які апробовано під час створення гнучких добре масштабованих мереж оператора мобільного зв'язку із повним або частковим застосуванням принципу віртуалізації мережевих функцій, зокрема під час оптимізації роботи мереж телекомунікаційної компанії Lifecell, що дозволило зменшити коефіцієнт простою обчислювальних ресурсів на 15%, уникнути зниження якості обслуговування під час сплесків перевантаження, утримати показники якості обслуговування на заданому рівні.

2. Запропоновано модифікації компонентів підсистеми PCRF (Policy and Charging Rules Function) та нові процедури організації їх взаємодії між собою та з підсистемами керування віртуалізованим середовищем для забезпечення контролю якості обслуговування гібридних телекомунікаційних сервісів, що дозволило запобігти погіршенню показників QoS на рівні абонентського пристрою кінцевого користувача та зменшити кількість відмов за рахунок скорочення часу реагування підсистеми керування телекомунікаційною та хмарною частинами гетерогенного середовища до 10%.

3. Запропоновану архітектуру телекомунікаційної мережі з віртуалізацією мережевих функцій із забезпеченням контролю якості обслуговування гібридних телекомунікаційних сервісів та керуванням обчислювальними ресурсами ТКС апробовано під час побудови сучасних систем типу SmartCity, M2M, D2D в лабораторіях компанії “Huawei Ukraine”, що дозволило ефективно перерозподіляти навантаження між групою фізичних та віртуальних вузлів обслуговування та покращити коефіцієнт використання обчислювальних ресурсів, а саме зменшити відсоток запитів, які було втрачено через перевищення допустимого часу обслуговування на 5%, зменшити відсотки часу низької завантаженості обчислювальних ресурсів на 8%, а відповідно їх перевантаження на 10%.

4. Запропоновані нові процедури та алгоритми: алгоритм планування обсягу ресурсів ГТС, процедура організації взаємодії підсистем ТКМ та підсистем керування віртуалізованим телекомунікаційним середовищем з забезпеченням контролю якості обслуговування, планування обсягу необхідних обчислювальних ресурсів для ефективної роботи ГТС, які використано в науково-дослідних роботах з розробки перспективних телекомунікаційних систем НДІ телекомунікацій та навчальному процесі кафедри ІТМ Інституту телекомунікаційних систем «КПІ ім. Ігоря Сікорського», результати моделювання запропонованих технічних рішень показали збільшення ймовірності утримання ресурсів обслуговування у заданих межах до 32%.

Особистий внесок здобувача. Усі наукові результати, що виносяться на захист дисертації, отримані здобувачем самостійно. Автору належить постановка задач досліджень, їх забезпечення, теоретичне обґрунтування, експериментальна перевірка нових моделей, методів та принципів.

У спільних публікаціях за темою дисертації автор зробив внесок, який полягає у формалізації ідей, виборі підходів до реалізації аналізу та узагальненні одержаних результатів. В роботах [1, 7] здобувачем систематизовано методи підвищення якості функціонування гетерогенної

телекомунікаційної системи та запропоновано концепцію диференційованого обслуговування у вузлі на границі локальної мережі. В роботах [16-17, 21] здобувачем запропоновано системний підхід до керування системою обслуговування мобільних абонентів, проаналізовано основні проблеми контролю якості комплексної мережі, які пов'язані з залученням додаткових ресурсів для обслуговування абонентів, запропоновано метод контролю якості на основі технології Smart Grid. В роботі [20] здобувачем запропоновано модель керування інфраструктурою NFV, яка враховує особливості розташування дата центрів обслуговування, а також структуру потоків гібридних телекомунікаційних сервісів, які поступають на обслуговування у дата центри.

В [12, 13] автором запропоновано принцип розділення функцій мережі зв'язку між телекомунікаційною та обчислювальною підсистемами, процедури їх взаємодії, а також модель процесу обслуговування гібридних телекомунікаційних сервісів послідовністю мережевих функцій.

В [5] автором запропоновано метод розподілу ресурсів обслуговування пропорційно вимогам сервісів різних типів для забезпечення показників якості обслуговування на заданому рівні В [15, 23] автором запропоновано спосіб керування навантаженням на гетерогенну телекомунікаційну систему, а також математичну модель і спосіб розрахунку оптимального навантаження на сервера системи онлайн тарифікації.

В [22] автором запропоновано спосіб резервування ресурсів у віртуалізованому середовищі для обслуговування надмірного потоку навантаження на елементи гетерогенного телекомунікаційного середовища, а у [14] метод розрахунку необхідних додаткових ресурсів.

В [6, 19] автором розроблено метод динамічного контролю використання ресурсів для забезпечення обслуговування вхідного потоку, критерії оцінки ефективності процесу обслуговування у складному ГТС. В [3] автором розроблено модель аналізу статистичних даних для формування прогнозу необхідних ресурсів на наступний період часу. В [2, 25] автором

запропоновано принцип прогнозування необхідного віртуального ресурсу у хмарі, спосіб моніторингу та прогнозування навантаження для забезпечення роботи ГТС. В [26] систематизовано математичні методи, які використовуються при проектуванні та оптимізації сучасних телекомунікаційних систем.

Апробація результатів дисертації. Основні положення і результати дисертаційної роботи були представлені, повідомлені й одержали схвалення на Науково-технічних конференціях "Проблеми телекомунікацій" (м. Київ, 2012-2015, 2017pp), IV Міжнародній конференції молодих вчених «Електроніка-2013», VII Міжнародній молодіжній науково-технічній конференції «РТ-2011» (Севастополь 2011, 2013 р), 20 – 24 Міжнародній конференції «СВЧ-техника и телекоммуникационные технологии. Крымико» (м. Севастополь, 2010-2014гг.), International Conference TCSET'2014 (Львів Славсько), 13 і 14 Міжнародній конференції CADSM 2015, 2017 (Lviv–Polyana), IEEE International Black Sea Conference on Communications and Networking 2016, Radio Electronics & Info Communications (UkrMiCo), International Conference. – IEEE, 2016, 2017гг, International Scientific and Practical Conference «Problems of Infocommunications. Science and Technology» (PICS&T-2017), X International symposium on Telecommunications – BIINTEL 2014, Sarajevo, Bosnia and Herzegovina, International Summer School on Latency Control for Internet of Services, 2017, KARLSTAD UNIVERSITY, ШВЕЦІЯ.

Публікації. Основні положення дисертації, які в достатній мірі висвітлюють результати роботи, що виносяться на захист, опубліковані у 68 наукових працях, у тому числі у 1 монографії, 25 статтях у наукових фахових виданнях (з них 2 статті у виданнях іноземних держав, 5 у виданнях України, які включені до міжнародних наукометричних баз), 2 патентах на корисну модель, 36 тезах доповідей в збірниках матеріалів конференцій, 4 статтях у інших виданнях.

Структура і обсяг дисертації. Дисертація складається із анотацій, вступу, 6 розділів основного змісту, висновків, списку використаних джерел і

2 додатків. Загальний обсяг роботи становить 416 сторінок друкарського тексту, в тому числі список літератури із 219 найменувань і два додатки на 30 сторінках, робота містить ілюстрації та таблиці.

РОЗДІЛ 1

АНАЛІЗ СУЧАСНОГО СТАНУ ПРОБЛЕМ КЕРУВАННЯ ІНФОРМАЦІЙНО-ТЕЛЕКОМУНІКАЦІЙНИМИ СИСТЕМАМИ

На сьогоднішній день спостерігається стрімкий розвиток інформаційно-телекомунікаційних систем, які використовують для надання різномірних сервісів абонентам та пристроям. За останні 10 років суттєво змінилися підходи та концепції організації систем зв'язку. З метою розширення можливостей телекомунікаційних систем, побудованих на основі спеціалізованого апаратного забезпечення, в умовах зміни підходів до організації обчислювальних процесів та інтеграції обчислювальних систем з телекомунікаційними (хмарні обчислення) світова наукова спільнота активно займається розробкою методологічної бази та стандартизацією роботи нових інформаційно-телекомунікаційних систем.

Розвиток систем зв'язку зумовив появу нових складових інформаційно-комунікаційних систем, означення яких наведено у сучасних стандартах та специфікаціях:

NFV - це концепція мережевої архітектури, яка пропонує використовувати технології віртуалізації для представлення цілих класів функцій мережевих вузлів у вигляді складових елементів, які можуть бути об'єднані разом або пов'язаних у ланцюг для створення телекомунікаційних послуг (сервісів).

Гібридний телекомунікаційний сервіс [1] - сервіс, який складається з компонентів телекомунікацій та хмарних сервісів.

Хмарні обчислення (Cloud Computing) [2] – парадигма забезпечення мережевого доступу до масштабованого і гнучкого набору спільно використовуваних фізичних або віртуальних ресурсів з наданням та адмініструванням ресурсів на основі самообслуговування за запитом.

Вузол обслуговування (відповідно до [3]) - мережева точка присутності (N-PoP) - місце, де мережева функція реалізується як Фізична мережева функція (PNF) або Функція віртуальної мережі (VNF).

Мережева функція (NF) [4] - це функціональний блок всередині мережевої інфраструктури, який має чітко визначені зовнішні інтерфейси та чітко визначену функціональну поведінку. На практиці мережева функція є мережевим вузлом або фізичним пристроєм.

Віртуалізовані мережеві функції (VNF) [4]: принцип розділення мережевих функцій на апаратне забезпечення, яке вони запускають, використовуючи абстракцію віртуальних апаратних засобів

Інфраструктура NFV [4]: сукупність всіх апаратних і програмних компонентів, які створюють середовище, в якому розміщені віртуалізовані мережеві функції (VNF)

NFV-Resource (NFV-Res) [4] існують всередині інфраструктури NFV та можуть бути використані для реалізації мережевих функцій або систем функцій, для забезпечення їхньої якості. До них можна віднести VeCPU – віртуальний процесор - ресурс віртуальної машини.

Компоненти інфраструктури NFV[4]: ресурси обладнання, які не є реплікабельними, але сприймаються як фізичні компоненти виготовлені в умовах виробництва.

Віртуальна машина (VM) [4]: віртуальний обчислювальний простір, який характеризується багатьма фізичними компонентами (процесор, пам'ять/сховище даних, порти/інтерфейси) та генерується за допомогою гіпервізору, який використовує фізичні ресурси та розміщує на них VM. VM можуть виступати хостінгом для окремих компонентів віртуальних машин.

Віртуальна мережа (virtual network VN) [4]: віртуальна мережа містить інформацію про маршрути віртуальних мережевих інтерфейсів VM та фізичних мережевих інтерфейсів, створюючи необхідну зв'язність. Віртуальна мережа обмежена набором доступних мережевих інтерфейсів.

Прикладний програмний компонент [4]: більш загальний термін для частини програмного забезпечення, яке може бути завантажене на віртуальну машину. Віртуальна мережева функція (VNF) є прикладом віртуального прикладного програмного компоненту.

Багатохмарні обчислення [9] – парадигма для взаємодії між двома та більше постачальниками хмарних послуг.

Для уточненого опису середовища, де здійснюється обслуговування гібридних телекомунікаційних сервісів, наведемо визначення гетерогенного телекомунікаційного середовища, яке є організаційно-технічною сукупністю, що складається з каналів зв'язку (фізичних та віртуальних), мережевих вузлів (фізичних та віртуальних) та обчислювального середовища, яке організовано відповідно до парадигми хмарних обчислень, забезпечує функціонування віртуалізованих телекомунікаційних сутностей.

Робота телекомунікаційних систем, які надають послуги зв'язку, сьогодні пов'язана з такими факторами та тенденціями:

1. Зростання об'ємів трафіків в експоненційній прогресії з кожним роком.
2. Мультисервісний потік потребує диференційованого обслуговування, різні вимоги до показників якості обслуговування, різні вимоги до організації процесу передачі, обліку та тарифікації трафіку.
3. Для гарантій заявлених показників якості обслуговування оператору необхідно постійно контролювати показники якості обслуговування абонентського сервісу та вчасно реагувати на зниження показників якості.

Саме тому для підтримки гнучкості телекомунікаційної системи оператору необхідно впроваджувати рішення з використанням програмного забезпечення на різних ділянках системи обслуговування.

Відповідно до рекомендацій ITU-T у складних об'єднаних послугах з багатим мультимедійним наповненням використовують різноманітні види інфраструктури електрозв'язку та інформаційних технологій (ІТ). Такі

послуги складаються з компонентів окремих послуг, придбаних у третіх сторін або наданим третім сторонам.

В рекомендації ITU-T Y.3520 визначено необхідність розробки рішень [4], які б забезпечили можливість однакового наскрізного керування (включаючи облік) послугами, які надаються окремо і спільно доменами і платформами різних постачальників хмарних послуг.

Досягти постійного забезпечення хмарними послугами, які належать до різних доменів – це нелегке завдання. Для його вирішення потрібен підхід, який би створював умови і підтримував постійний керований доступ до хмарних послуг.

Для досягнення зазначених вище цілей постачальникам хмарних послуг потрібні структурні основи, архітектура, шаблони проектування та приклади передового досвіду [5].

Для забезпечення можливості швидкої розробки і розгортання складових багатохмарних послуг в галузі електрозв'язку, ефективного надання сучасних послуг кінцевим користувачам, оперативного керування ними необхідні відповідні удосконалені стандартні моделі та методи керування процесом обслуговування у гетерогенному телекомунікаційному середовищі.

1.1. Мережі NGN як основа для 5G. Особливості системи керування.

Згідно з визначенням, наведеним в Рекомендації MCE-T Y.2001, мережа наступного покоління (NGN) - це мережа з пакетною комутацією, здатна забезпечити користувачів різноманітними вузькосмуговими і широкосмуговими послугами, включаючи послуги телефонного зв'язку. Вона заснована на широкосмуговій мережі з пакетною технологією транспортування, що забезпечує необхідну якість послуг QoS, в якій функції, пов'язані з наданням послуг, не залежать від технологій транспортування інформації. Мережа NGN дає користувачам необмежений доступ до різноманітних послуг провайдерів і підтримує узагальнену мобільність, яка

дозволяє користувачам отримати доступ до послуг у будь-якому місці і в будь-який час.

У рекомендації MCE-T Y.2012 перераховані основні принципи функціональної архітектури NGN:

Підтримка багатьох технологій доступу - функціональна архітектура NGN вимагає гнучкої конфігурації, необхідної для підтримки груп технологій доступу.

Розподілене керування - використання принципу розподіленої обробки в пакетних мережах і підтримка прозорості розташування для розподілених обчислень.

Відкрите керування - мережеві інтерфейси керування мають бути відкритими для підтримки процесів створення нових, зміни існуючих послуг та підтримки засобів забезпечення логіки послуг сторонніх постачальників.

Незалежність надання послуг - процес надання послуг має бути розділений між функціями транспортної мережі, яка працює з використанням механізму розподіленого відкритого керування. Це підтримує конкурентне оточення під час розвитку NGN, сприяє прискоренню процесів впровадження нових послуг.

Для реалізації цих функцій в Рекомендації ITU-T Y.2011 [5] запропоновано базову еталонну модель NGN, яка включає два рівні: рівень послуг NGN (service stratum) і рівень транспорту NGN (transport stratum), кожен з яких містить по три площини: користувача, керування та менеджменту (Рис. 1.1).

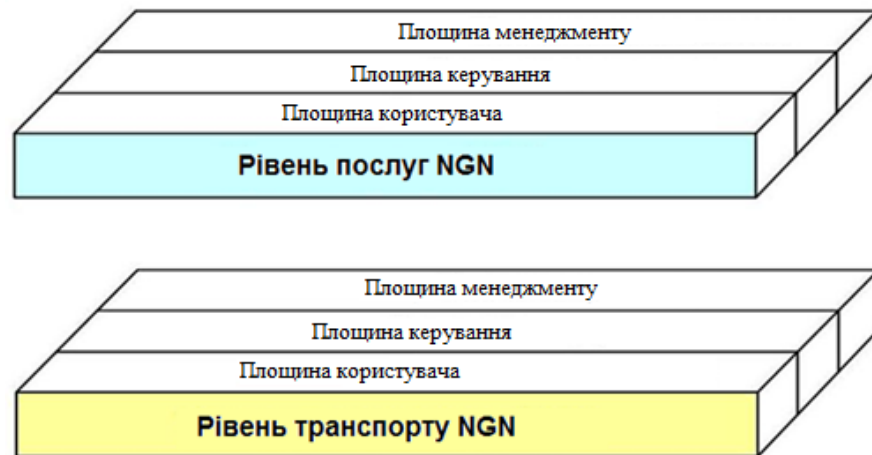


Рис. 1.1 Базова еталонна модель NGN (із рекомендації МСЕ-Т Y.2011)

Функціональність рівнів базової еталонної моделі NGN наведено на Рис. 1.2 [6] (рекомендації ITU-T Y.2012). На кожному з рівнів використовують декілька функцій, так для надання послуг (прикладних сервісів кінцевим користувачам) застосовують функції підтримки прикладних сервісів і послуг, відповідні керуючі функції. NGN підтримує точку сполучення з функціональною групою прикладних сервісів, так званий інтерфейс прикладних сервісів мережі (application network interface-ANI), який реалізує канал взаємодії та обміну інформацією між прикладними сервісами і елементами мережі NGN. ANI забезпечує ресурси, необхідні для реалізації прикладних сервісів. Транспортний рівень забезпечує послуги IP-з'єднань для користувачів мережі NGN за допомогою функцій керування транспортом, включаючи функції керування мережевими підключеннями NACFs (Network Attachment Control Functions) і функції керування ресурсами та доступом RACFs.

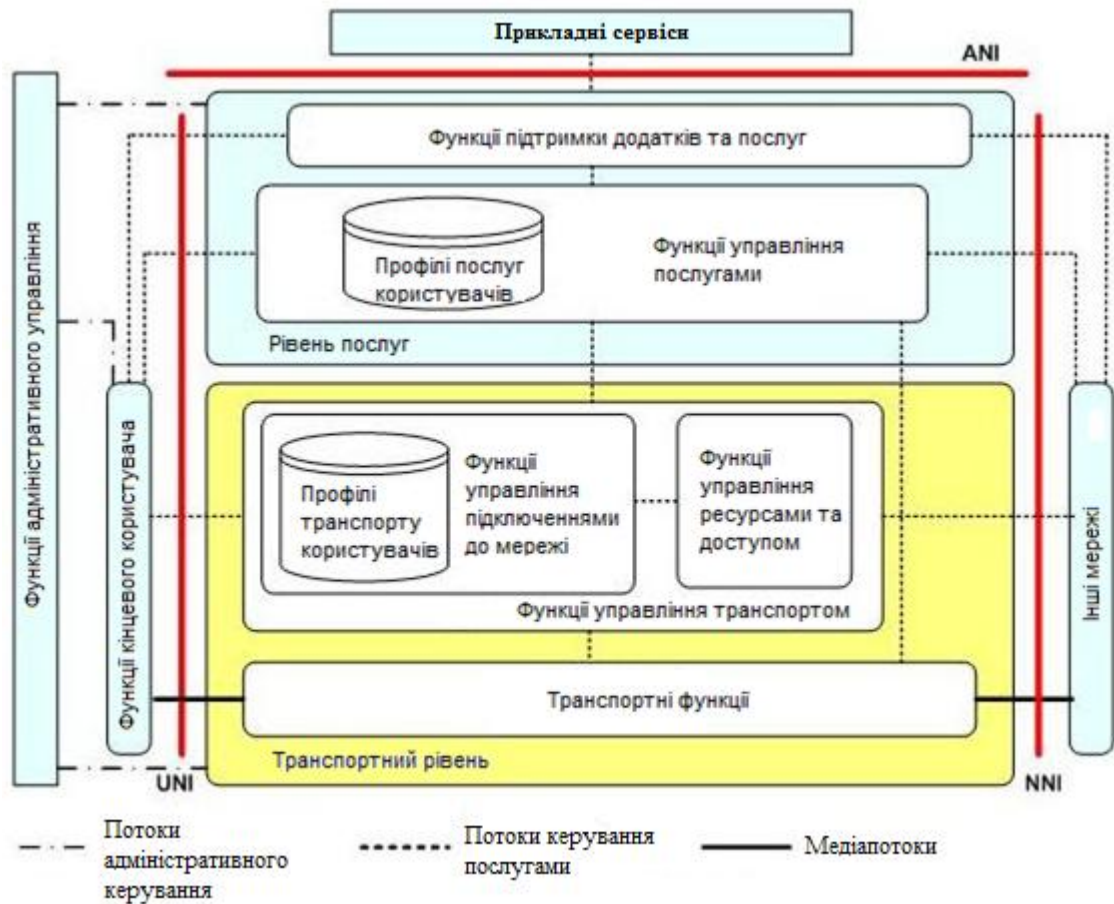


Рис. 1.2 Загальна функціональна архітектура NGN

Відповідно до Рекомендації МСЕ-Т У.2011 функції транспортного рівня включають безпосередньо транспортні функції і функції керування транспортом.

Транспортні функції (transport functions) забезпечують з'єднання всіх компонент і фізично розділених функцій всередині NGN. Ці функції підтримують передачу медіаінформації, а також інформації керування (сигналізації) та технічного обслуговування. Транспортні функції включають функції мережі доступу, прикордонні функції, функції транспортного ядра (магістралі) і функції шлюзів.

Функції мережі доступу (access network functions) забезпечують підключення кінцевих користувачів до мережі, а також збір і агрегацію трафіку, що надходить з мережі доступу в транспортну магістраль (ядро). Ці функції також реалізують механізми керування якістю обслуговування QoS, пов'язані безпосередньо з трафіком користувача, включаючи керування

буферами, чергами і розкладами, пакетну фільтрацію, класифікацію трафіку, маркування трафіку, визначення політик обслуговування і формування профілю передачі трафіку.

Функції мережі доступу залежать від використовуваної технології доступу, наприклад, вони відрізняються для бездротової технології CDMA та провідної технології доступу xDSL. Залежно від технології, яка використовується для доступу до послуг NGN, мережа доступу включає функції, пов'язані з:

- кабельним доступом;
- доступом за технологіями xDSL;
- бездротовим доступом (наприклад, технології IEEE 802.11 (WiFi), 802.16 (WiMAX), доступ 3G RAN);
- оптичним доступом.

Прикордонні функції (edge functions) використовують для обробки трафіку, який виходить шляхом агрегування трафіку, що надходить з різних мереж доступу і передається в магістральну транспортну мережу. Вони включають функції, пов'язані з підтримкою якості обслуговування QoS і керування трафіком. Прикордонні функції використовують також між магістральними транспортними мережами.

Магістральні транспортні функції (core transport functions) відповідають за гарантовану передачу інформації через транспортну мережу з різним рівнем якості. Вони забезпечують механізми реалізації заданого рівня якості передачі QoS для користувача трафіку включаючи керування буферами, чергами і розкладом, фільтрацію пакетів, класифікацію, маркування і формування трафіку, контроль дотримання правил обслуговування, керування шлюзами і функції міжмережевих екранів.

Функції шлюзів (gateway functions) забезпечують можливості взаємодії з функціями кінцевих користувачів і/або іншими мережами, включаючи інші типи мереж NGN та ряд існуючих мереж, таких як ТфЗК/ISDN, публічний

Інтернет та інші. Функції шлюзів можуть керуватися або безпосередньо функціями рівня керування або через функції керування транспортною мережею.

Функції обробки медіаінформації (media handling functions) забезпечують обробку медіаінформації при наданні послуг, таких як генерація тональних сигналів і перекодування. Ці функції реалізують спеціальними ресурсами обробки медіаінформації на транспортному рівні.

Функції керування транспортною мережею (transport control functions) включають функції керування ресурсами та доступом, функції керування приєднанням до мережі.

Функції керування ресурсами та доступом RACFs діють як арбітр між функціями керування послугами і транспортними функціями для підтримки QoS і пов'язані з керуванням транспортними ресурсами в мережі доступу і в магістральній транспортній мережі. Рішення з керування ґрунтуються на інформації про необхідний транспорт, угодах про заданий рівень обслуговування SLA, правилах мережевої політики, пріоритетах послуг та інформації про стан і використання транспортних ресурсів. Функції RACF забезпечують абстрактний підхід до інфраструктури транспортної мережі для функцій керування послугами SCFs і надають сервіс-провайдерам можливість не залежати від мережевої топології, зв'язності, завантаження ресурсів, механізмів/технологій QoS та ін. Функції RACF взаємодіють з функціями SCF і транспортними функціями для різних прикладних програмних компонентів (наприклад, SIP-виклики, потокове відео й ін.), що вимагає керування транспортними ресурсами NGN, включаючи керування QoS, керування NAT/Firewall і проходження трансляції мережевих адрес на рівні портів NAT.

Функції керування підключенням до мережі NACFs забезпечують реєстрацію на рівні доступу та ініціалізацію функцій кінцевого користувача для послуг доступу NGN. Ці функції забезпечують транспортний рівень ідентифікацією/авторизацією, керуючи простором IP-адрес в мережі доступу

і аутентифікації сесій доступу. Вони також повідомляють кінцевим користувачам про контактні точки до функцій NGN на рівні послуг. Функції NACF включають транспортний профіль користувача, який зберігатися у вигляді функціональної бази даних, що включає інформацію користувача, а також інші дані керування.

Рівень послуг (service stratum) включає:

- Функції керування послугами, включаючи функції профілів послуг користувачів;
- Функції підтримки прикладних програмних компонентів і функції підтримки послуг.
- Функції кінцевих користувачів.
- Функції адміністративного керування

Функції керування послугами (service control functions) включають керування ресурсами, функції реєстрації, аутентифікації та авторизації для різних сервісів на рівні послуг. Вони також можуть включати функції керування медіаресурсами, такими як спеціалізовані пристрої та шлюзи на сигнальному рівні. Функції керування послугами підтримують профілі послуг користувачів, які є комбінацію інформації користувача та інших даних керування, індивідуальний профіль кожного користувача, такі дані зберігають у функціональних базах даних.

Функції підтримки прикладних програмних компонентів і функції підтримки послуг (application support functions and service support functions) включають функції шлюзів, реєстрації, аутентифікації та авторизації на рівні прикладних програмних компонентів. Ці функції доступні в функціональних групах «прикладні сервіси» і «кінцеві користувачі». Вони працюють спільно з функціями керування послугами для забезпечення кінцевих користувачів і прикладних сервісів необхідними послугами NGN. Через інтерфейс «користувач-мережа» UNI функції підтримки прикладних програмних компонентів і функції підтримки послуг забезпечують точку доступу до

функцій кінцевих користувачів. Взаємодія прикладних програмних компонентів з даними функціями здійснюється через точку доступу, реалізовану інтерфейсом «прикладна програма-мережа» ANI.

Функції кінцевих користувачів (end-user functions) не визначають ніяких обмежень на інтерфейси користувача і мережі доступу кінцевих користувачів, які можуть бути з'єднані з мережею доступу NGN. Термінальні пристрої користувачів послуг NGN є будь-якими мобільними або стаціонарними пристроями.

Функції адміністративного керування (management functions) забезпечують можливість керувати мережею NGN для надання послуг із заданим рівнем якості, безпеки та надійності. Ці функції розподіляються децентралізовано по всім функціональним блокам (FE) і вони взаємодіють з функціональними блоками керування мережевими елементами, керування мережею і керування послугами. Функції адміністративного керування використовують на транспортному рівні і рівні послуг і для кожного з рівнів вони реалізують такі завдання:

- Керування процесом усунення відмов (Fault Management);
- Керування конфігурацією мережі (Configuration Management);
- Керування розрахунками з користувачами і постачальниками послуг (Accounting Management);
- Контроль продуктивності мережі (Performance Management);
- Забезпечення безпеки роботи мережі (Security Management).

З метою більш простого розуміння принципів побудови мереж наступного покоління в більшості публікацій з NGN наводиться узагальнена 4-х рівнева архітектура NGN, в якій виділяються такі рівні (Рис. 1.3):



Рис. 1.3 Чотирьохшарова модель NGN

- рівень доступу, який містить мережу абонентського доступу до транспортної пакетної мережі;
- транспортний рівень, який включає магістральну пакетну мережу (мережу, побудовану на базі протоколів пакетної комутації IP або ATM, на сьогоднішній день найчастіше на базі технології MPLS та протоколу IP);
- рівень керування комутацією, включає сукупність функцій з керування усіма процесами обслуговування викликами в телекомунікаційній мережі;
- рівень послуг та експлуатаційного керування, який містить логіку виконання послуг та/або прикладних програмних компонентів, керує цими послугами, має відкриті інтерфейси для використання сторонніми організаціями (для розробки нових сервісів).

Термінальне обладнання рівня доступу не входить до складу мережі NGN. Безпосереднє підключення до мережі можливо тільки для пакетних абонентських терміналів, які працюють з використанням протоколів SIP та H.323.

Концепція NGN спричинила еволюцію систем керування в телекомунікаціях від спеціальних апаратних рішень до програмно-керованих

мереж (SDN), які здійснюють контроль та керування відповідним обладнанням гетерогенних інформаційно-телекомунікаційних систем.

1.2. Система організації хмарних обчислень. Аналіз стандартів для забезпечення керування телекомунікаційними системами

Хмарні обчислення (Cloud Computing) [ITU-T Y.3500]: Парадигма забезпечення мережевого доступу до масштабованого і гнучкого набору спільно використовуваних фізичних або віртуальних ресурсів з наданням та адмініструванням ресурсів на основі самообслуговування за запитом.

Прикладами таких ресурсів є сервери, операційні системи, мережі, програмне забезпечення, програми та обладнання для зберігання даних.

Договір про рівень обслуговування (Service Level Agreement) [ITU-T Y.3500]: Документально оформлена угода між постачальником послуги і споживачем, в якій визначаються послуги та цільові показники обслуговування.

Угода про рівень обслуговування може укладатися також між провайдером послуги і постачальником, внутрішньої групою або споживачем, який виступає в ролі постачальника. Договір про рівень обслуговування може включатись в договір або в документально оформлену угоду іншого типу.

Керування ресурсами (Resource management) [5]: Найбільш економічний і ефективний спосіб доступу, контролю, керування, розгортання, планування і зв'язування ресурсів, які надаються постачальниками послуг.

Хмарні прикладні програмні компоненти, які також називаються хмарними робочими навантаженнями, – це прикладні програмні компоненти (тобто програми конкретного цільового призначення), які повинні виконуватися в центрах обробки даних постачальника хмарних послуг для створення екземплярів хмарних послуг і забезпечення їх доступності користувачам. Іншими словами, хмарний прикладний програмний компонент

повинен виконуватися для того, щоб надати користувачам одну або кілька хмарних послуг.

Постачальникам хмарних послуг необхідно виробити глибоке розуміння аспектів надання послуг, які відносяться до часу виконання, а також методів керування цими послугами і ресурсами, необхідними для їх надання [5].

У складних об'єднаних послугах з багатим мультимедійним наповненням використовують різноманітні види інфраструктури електрозв'язку та інформаційних технологій (ІТ). Такі послуги складаються з компонентів окремих послуг, придбаних у третіх сторін або наданим третім сторонам

Одна з цілей рекомендації ITU-T Y.3520[5] полягає в тому, щоб забезпечити можливість однакового наскрізного керування (включаючи облік) послугами, які надаються окремо і спільно доменами і платформами різних постачальників хмарних послуг.

Досягти постійного забезпечення хмарними послугами, які належать до різних доменів – це нелегке завдання. Для його вирішення бажано виробити підхід, який би створював умови і підтримував постійний керуванський доступ до хмарних послуг.

Для досягнення зазначених вище цілей постачальникам хмарних послуг потрібні структурні основи, архітектура, шаблони проектування та приклади передового досвіду [5]. Інтерфейси компонентів окремих послуг не є основним предметом розгляду даної роботи, оскільки фізичні інтерфейси можуть відрізнятися в залежності від реалізації, використовуваних сторонніх технологій і вимог операторів. Для забезпечення можливості швидкої розробки і розгортання складових багатохмарних послуг в галузі електрозв'язку, а також оперативного керування ними, необхідні відповідні стандартні принципи проектування та базові структури.

1.2.1. Особливості керування сервісами із застосуванням хмарних обчислень

Підвищена складність керування ресурсами, пов'язаними з наданням хмарних послуг, обумовлена двома основними особливостями керування сервісами із застосуванням хмарних обчислень. Перша особливість – це віртуалізація обчислювальних і мережних ресурсів в еталонній архітектурі хмарних обчислень [7]. Друга особливість полягає в тому, що в наданні хмарних послуг все частіше беруть участь домени великої кількості постачальників хмарних послуг, і таке гетерогенне середовище надзвичайно ускладнює наскрізне керування ресурсами.

В цілому керування ресурсами слід розглядати з точки зору керування життєвим циклом хмарного прикладного програмного сервісу. На різних етапах свого життєвого циклу він має піддаватися впливам з боку традиційних бізнес-процесів, пов'язаних з виконанням функцій системи керування, таких як адміністрування, підготовка, конфігурація, забезпечення якості послуг і нарахування плати.



Рис. 1.4 Прикладні програмні сервіси, розміщені в однохмарній обчислювальній системі

Як показано на Рис. 1.4, в простому випадку, коли прикладний програмний сервіс розміщений в однохмарній обчислювальній системі, вона стає залежною від двох окремих категорій віртуалізованих ресурсів.

Пунктирними стрілками позначений активний координований зв'язок, який необхідно підтримувати між ресурсами на кожному рівні.

Хоча на рис. 1.4 віртуалізовані ресурси діляться на хмарні і мережеві, в хмарних обчисленнях вважається, що всі ресурси розташовуються на одному рівні [4].

Необхідно вирішити питання про те, як засобами існуючих систем керування хмарою забезпечити відстеження логічних і фізичних ресурсів, що реально підходять для певного екземпляру конкретного прикладного програмного сервісу в будь-який заданий момент часу.

Передові практичні методи мають забезпечувати гнучкість хмарного прикладного програмного сервісу з точки зору надання доступу до своїх інтерфейсів керування послугами або до ресурсів. Крім того, вони мають передбачати можливість відкриття доступу до одного або декількох інтерфейсів системи керування, щоб та відстежувала динамічні зміни основних ресурсів, виділених для підтримки роботи керованого хмарного прикладного програмного компонента.

На Рис.1.5 показана високорівнева архітектура наскрізного керування багатохмарними ресурсами. Тут зображені віртуальні машини з програмним стеком, який складається з міжплатформених рівнів з серверами прикладних програмних компонентів, що працюють в обраному середовищі виконання, поверх яких виконуються хмарні прикладні програмні компоненти.

Крім того, на Рис.1.5 представлені функціональні інтерфейси (FI) і інтерфейси керування послугами (SMI), до яких надають доступ різні хмарні прикладні програмні компоненти, що працюють в центрах обробки даних від безлічі хмар. Необхідну інформацію можна збирати через будь-які інтерфейси SMI, пов'язані з великою кількістю прикладних програмних компонентів, що виконуються в центрах обробки даних від безлічі хмарних послуг. Це дозволяє реалізувати всеосяжне наскрізне керування багатохмарними ресурсами і систему моніторингу.

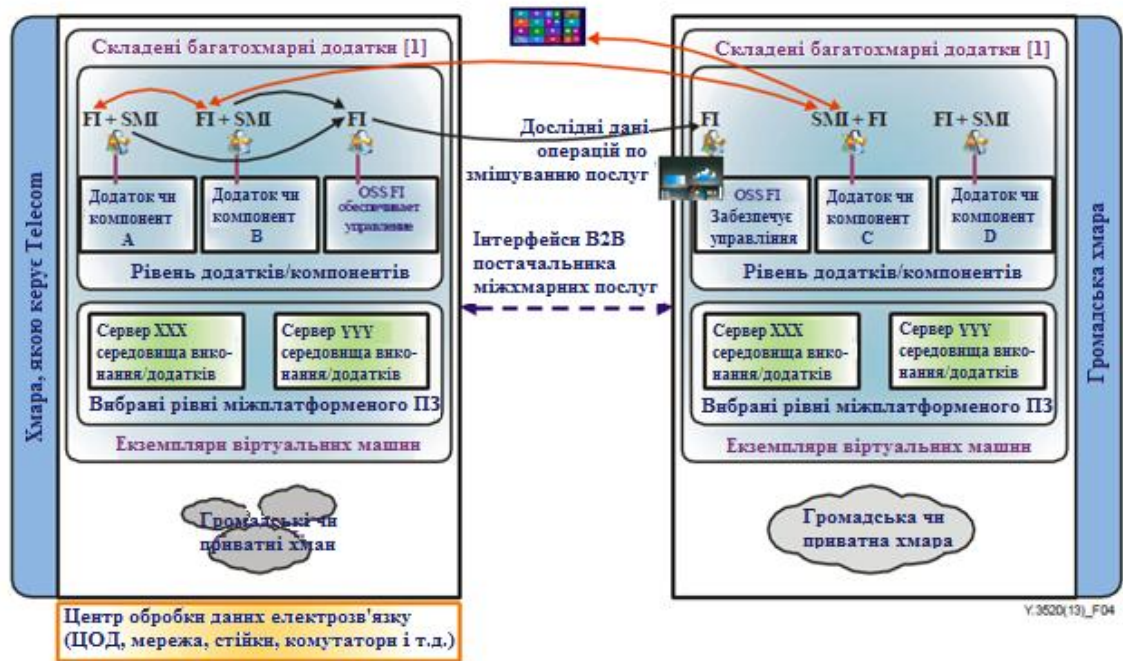


Рис.1.5 Концептуальна архітектура багатохмарного і багатоплатформеного керування хмарними послугами

На Рис.1.5 прикладні програмні компоненти, які працюють у віртуальних машинах, можуть бути об'єднаними, розподіленими, побудованими з різних програмних компонентів. Окремий екземпляр віртуальної машини може містити всі програмні компоненти, що відносяться до такого прикладного програмного компонента, або ж тільки деякі з них у разі розподіленого прикладного програмного компонента, що виконується в декількох віртуальних машинах (звідси посилання на прикладні програми або компоненти на Рис.1.5).

Концептуальна архітектура, наведена на Рис.1.5, дозволяє створювати функціонально сумісні програми з підтримкою сценаріїв виносу в хмару або гібридних хмарних обчислень.

Між-хмарне обчислення (inter-cloud computing) [8]: парадигма для взаємодії між двома або більше постачальниками хмарних послуг.

Телекомунікаційний сервіс (telecommunication service) [9]: те, що адміністрація пропонує своїм клієнтам для задоволення конкретних вимог до електрозв'язку.

Сервіс передачі (bearer service) та телесервіс (teleservice) є різновидами служби телекомунікацій. В майбутньому можуть бути визначені інші види телекомунікаційних послуг.

Інтерфейс керування сервісом (service management interface) [МСЕ-Т М.3070]: інтерфейс, який надає набір можливостей керування, що виникають в рамках хмарного сервісу. З його допомогою здійснюється керування хмарним сервісом.

Постачальник послуг (service provider) [b-ITU-T М.3320]: загальне визначення суб'єкта, який надає телекомунікаційні послуги клієнтам та іншим користувачам на тарифній або контрактній основі. Постачальник послуг може здійснювати керування мережею. Постачальник послуг може бути замовником іншого постачальника послуг.

ТС-гібридний сервіс (ITU-T М.3371): сервіс, який складається як з компонентів телекомунікацій, так і хмарних сервісів.

Як важлива частина керування телекомунікаціями, керування сервісом реалізує всі функції, необхідні для роботи комунікаційних та інформаційних послуг, що надаються клієнтам. Сюди входить виконання послуг, надання послуг та їх оплата протягом всього життєвого циклу послуги:

- виконання послуг: виконання ресурсних можливостей та вимог до якості обслуговування клієнтів, підтримка готовності ресурсного та сервісного потенціалу, забезпечення функціональності та інтерфейсу, необхідного клієнтам для доступу та споживання послуг;
- забезпечення послуг: підтримка функціональності та процесу надання послуг включає в себе керування рівнем обслуговування, керування інцидентами та проблемами, сервісний моніторинг та керування звітністю;
- білінг послуг: підтримка функції та процесу оплати послуг.

У зв'язку з наближенням індустрії інформаційних та комунікаційних технологій, у сфері телекомунікаційної інфраструктури все більше використовують хмарні обчислення. Телекомунікаційні оператори надають хмарні сервіси, а також застосовують технології хмарних обчислень для

оптимізації своїх телекомунікаційних платформ та систем підтримки (ITU-T M.3371).

[ITU-T M.3070] визначає загальну модель, засновану на інтерфейсі керування послугами (SMI), для керування хмарними обчисленнями з кінця в кінець (end-to-end E2E), яка описана на Рис. 1.6. У цій моделі система керування телекомунікаціями, яка працює у хмарі, може керувати хмарними об'єктами через інтерфейс керування "I1", який може відповідати набору SMI.

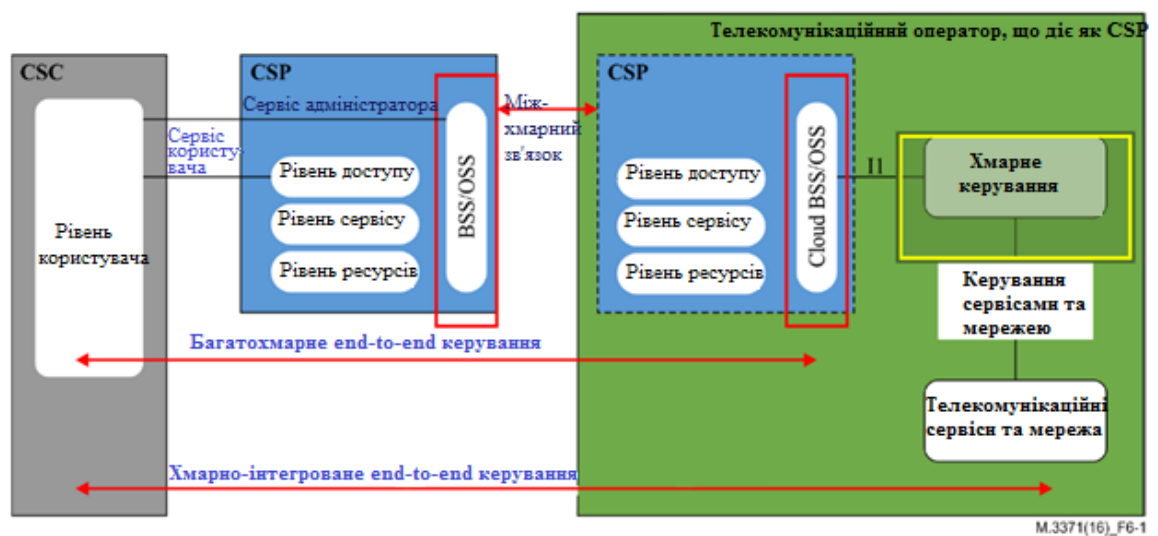


Рис. 1.6 Спільна модель керування хмарними обчисленнями E2E, з оператором зв'язку, що виступає в ролі CSP (після [MCE-T M.3070])

Система керування телекомунікаціями, що працює у хмарі, може послідовно керувати телекомунікаційними та хмарними сервісами. Незважаючи на те, що поділ на рівні керування може бути таким самим, як в класичному телекомунікаційному менеджменті, з впровадженням хмарних обчислень існують ще нові вимоги до керування хмарними телекомунікаціями. Ця рекомендація стосується вимог комплексного керування послугами в хмарній системі керування телекомунікаціями.

Співвідношення між керуванням хмарними сервісами та керуванням сервісами в хмарних телекомунікаціях також показано на Рис. 1.6. Керування хмарними службами знаходиться в системі підтримки бізнесу (BSS) /системі підтримки операцій (OSS) (позначено червоним прямокутником на Рис. 1.6),

що належить постачальнику хмарних сервісів (CSP) а керування службами у хмарних телекомунікаціях виконує інформаційна система керування телекомунікаціями (позначена жовтим прямокутником на Рис. 1.6), що належить оператору телекомунікацій. Ще одним важливим аспектом керування хмарними сервісами є керування життєвим циклом хмарних сервісів (див. [ITU-T Y.3522]), що також знаходиться в BSS/OSS, і належить CSP.

1.2.2. Загальні вимоги до керування сервісами у гетерогенному інформаційно-телекомунікаційному середовищі

Відмінність хмарної системи керування телекомунікаціями від традиційної полягає у підтримці комплексного керування телекомунікаційними сервісами, хмарними сервісами та ТС-гібридними сервісами, які одночасно складаються з телекомунікаційних і хмарних сервісів.

До хмарної системи керування телекомунікаціями висуваються такі вимоги:

- керування послугами повинно забезпечувати керування функціональністю каталогів сервісів та інвентаризації сервісів для забезпечення телекомунікаційних послуг, хмарного сервісу та ТС-гібридної служби;
- керування послугами повинно підтримувати модифікацію та припинення (на замовлення та автоматично) надання телекомунікаційних послуг, хмарного сервісу та ТС-гібридної послуги;
- керування сервісами повинно підтримувати забезпечення якості E2E у сфері телекомунікаційних послуг, хмарного сервісу та ТС-гібридної послуги, а також забезпечувати високий рівень надійності та доступності, відповідно до угоди про рівень обслуговування (SLA);

- керування сервісами повинно підтримувати тарифікацію телекомунікаційних послуг, хмарного сервісу та ТС-гібридних послуг, відповідно до часу використання, пропускнуої здатності, ресурсів та будь-якої їх комбінації;

- керування сервісами повинно забезпечувати моніторинг, аудит та надання звітності для телекомунікаційних послуг, хмарного сервісу та ТС-гібридних сервісів з метою оцінки та забезпечення якості надання послуг.

Всі перелічені вище вимоги відносяться як до послуг, що надаються єдиним CSP, так і до міжхмарного провайдера.

Функціональна програмна платформа для керування послугами в міжхмарній системі керування

Високорівнева організація функцій керування сервісами в хмарній системі керування телекомунікаціями складається з керування сервісним каталогом, інвентаризацією послуг (service inventory), тестуванням, замовленням, якістю надання послуг, ефективністю обслуговування та оцінки послуг/дисконтуванням.

На Рис. 1.7 зображено функціональну основу для керування сервісами в хмарній системі керування телекомунікаціями:

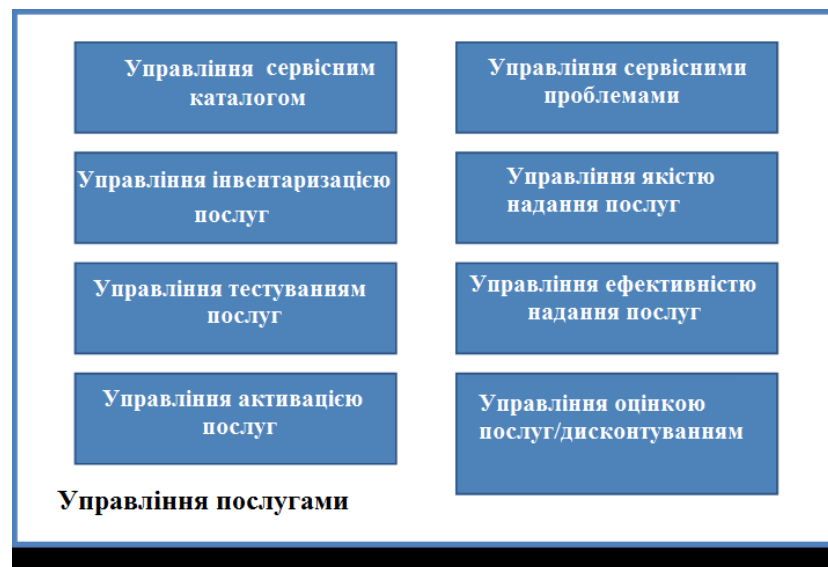


Рис. 1.7 Функціональна програмна платформа для керування послугами в міжхмарній системі керування

Функції наступні:

- керування каталогом сервісів: надає можливості для створення та розробки нових послуг, визначення їх характеристик, керування комплексними правилами, забезпечує підтримку компонування послуг та керування їх зв'язками і залежностями;

- керування інвентаризацією послуг: надає можливості для зберігання та керування елементами послуг та їх атрибутами. Функціональна платформа також зберігає й керує взаємодією сервісів, тобто відображенням послуг одних послуг для інших та/або їх компонентів;

- керування тестуванням послуг: надає можливості для забезпечення роботи служб належним чином. У процесі виконання, тестування служби відповідає за те, щоб зазначена служба працювала так, як повинна за замовчуванням. У процесі надання послуг, тестування відповідає за розпізнавання проблем та недоліків;

- керування активністю послуг: надає можливості для керування життєвим циклом E2E запиту на обслуговування. Вона охоплює перевірку наявності послуг, а також порядок надходження запитів на обслуговування. Сюди також входить порядок розбиття послуг та/або продуктів на частини, а також відстеження їх обслуговування, процесів активації та тестування;

- керування сервісними проблемами: надає можливості для керування проблемами, що впливають на роботу з клієнтом, а також мережевими проблемами/недоліками та їх ефективне вирішення;

- керування якістю послуг: надає можливості для моніторингу та керування рівнями обслуговування. Показники якості надання послуг збираються та порівнюються з еталонними, а висновки надаються зацікавленим сторонам;

- керування ефективністю обслуговування: надає можливості для моніторингу, аналізу та звітування про ефективність надання послуг E2E. Дана функція повинна давати можливості для аналізу E2E, щоб забезпечити правильне функціонування кожної служби, а також історичний огляд;

- керування оцінкою послуг/дисконтуванням: надає можливості для забезпечення отримання клієнтом рахунку, з відображенням всіх оплачуваних подій, що надаються постачальником послуг, відповідно до ділових взаємовідносин між ними.

1.2.3. Функціональні вимоги до керування сервісами в хмарній системі керування телекомунікаціями

В даному підрозділі наведені функціональні вимоги до керування сервісами в хмарній системі керування телекомунікаціями.

Керування каталогом послуг

У хмарній системі керування телекомунікаціями вимоги керування каталогом послуг наступні:

- керування каталогом послуг повинно надавати функції для керування інформацією про хмарні послуги. Потрібно, щоб керування каталогом послуг забезпечувало функції для керування інформацією ТС-гібридних послуг.

Керування інвентаризацією послуг

У хмарній системі керування телекомунікаціями вимоги до інвентаризації послуг наступні:

- керування інвентаризацією сервісів повинно забезпечувати функції для керування атрибутами хмарного сервісу;
- керування інвентаризацією сервісів повинно забезпечувати функції для відстеження взаємозв'язку між хмарним сервісом або гібридною службою ТС і ресурсами.

Керування активацією послуг

У хмарній системі керування телекомунікаціями вимоги до активації послуг наступні:

- надавати функції для перевірки наявності хмарних сервісів, а також запитів на замовлення послуг;

- надавати функції для керування розкладом замовлення послуг та продуктів;
- надавати функції для відстеження процесу підготовки активації послуги.

Керування сервісними проблемами

У хмарній системі керування телекомунікаціями вимоги щодо керування сервісним проблемами наступні:

- надавати функції для керування запитами про скарги з боку клієнта, а також повідомленнями від системи моніторингу про несправність мережі;
- надавати функції для розподілення помилок між хмарною та телекомунікаційною мережею;
- мати цілісне уявлення про конфігурацію та взаємозв'язок між різними мережевими рівнями для забезпечення аналізу первинної причини виникнення проблеми.

Керування ефективністю надання послуг

У хмарній системі керування телекомунікаціями вимоги щодо керування ефективністю надання послуг наступні:

- забезпечувати моніторинг, аналіз та надання звітності про ключовий показник ефективності (KPI) хмарного сервісу.

Керування тестуванням послуг

У хмарній системі керування телекомунікаціями вимоги щодо керування тестуванням послуг наступні:

- надавати функції для тестування хмарного сервісу як у процесі виконання, так і в процесі підтвердження надання послуги.

Керування якістю послуг

У хмарній системі керування телекомунікаціями вимоги щодо керування якістю надання послуг наступні:

- надавати функції моніторингу, аналізу та звітування про ключовий показник якості (KQI) хмарного сервісу.

Керування оцінкою послуг/дисконтуванням

У хмарній системі керування телекомунікаціями вимоги щодо керування оцінкою послуг/дисконтуванням наступні:

- надавати функції для підтримки оцінки та дисконтування послуг в контексті комплектування та складання сервісів із хмарним обслуговуванням.

Наприклад, за рахунок застосування в керуванні концепції загальних моделей мережі, стає можливим загальне керування різнотипним обладнанням, мережами і послугами, що використовують загальні інформаційні моделі і стандартні інтерфейси (ITU-T M.3060 / Y2401 (03/2006)).

Керування мережами електрозв'язку направлене на підтримку широкої різноманітності областей керування, що охоплює планування встановлення, експлуатацію, адміністрування, технічне обслуговування та надання мереж і послуг електрозв'язку.

В МСЕ-Т керування ділиться на п'ять широких функціональних областей керування (Рекомендація МСЕ-Т М.3400). На сьогоднішній день визначені наступні п'ять функціональних областей керування FCAPS:

- керування обробкою відмов;
- керування конфігурацією;
- керування обліком;
- керування якістю роботи;
- керування безпекою.

Справжня класифікація інформаційного обміну в структурі керування не залежить від того, яким чином буде використовуватися інформація.

Справжня класифікація інформаційного обміну в структурі керування не залежить від того, яким чином буде використовуватися інформація. Для керування мережами електрозв'язку мережі й послуги повинні розглядатися як сукупності взаємодіючих систем. Бізнес-процеси, описані в Рекомендаціях МСЕ-Т серії М.3050.x, і функціональні області керування FCAPS (Обробка відмов, конфігурація, облік, якість роботи, безпека), описані в Рекомендації

МСЕ -Т М.3400, повинні розглядатися як теоретичні побудови, необхідні для мереж і послуг наступних поколінь. Архітектура передбачає таку організацію керування окремими системами, при якій на мережі досягається узгоджений результат.

Завдання в галузі керування мережами наступних поколінь включають в себе:

- мінімізацію посередницької роботи між різними технологіями мереж шляхом зближення підходів до керування і використання інтелектуальної звітності;
- мінімізацію часу реагування системи керування на події в мережі;
- мінімізацію навантаження, створюваного трафіком керування;
- географічно розосереджений контроль над аспектами експлуатації мережі;
- надання механізмів розмежування для мінімізації ризиків у сфері безпеки;
- надання механізмів розмежування для виявлення і запобігання збоїв мережі;
- поліпшення допомоги в обслуговуванні і взаємодії з клієнтами;
- багаторівневе представлення послуг, що дозволяє постачальнику надавати структурні блоки для послуг, а також групувати послуги і їх вплив на архітектуру керування;
- бізнес-процеси, визначені в Рекомендаціях серії М.3050.х, а також їх майбутнє використання в СПП;
- підтримку прикладних програмних компонентів як на єдиній розподіленій обчислювальній платформі, так і розподілених по мережі.

Для подальших досліджень визначені наступні області:

- значення потреби в керуванні наскрізними послугами;
- значення домашніх мереж і обладнання в приміщенні користувача.

У той час як рівень керування обслуговуванням (РУО) відповідає за керування життєвим циклом послуг, а також за доставку і забезпечення

примірників послуг, рівень керування ресурсами (РУР) відповідає за керування логічного інфраструктурою обслуговування та логічною транспортною інфраструктурою.

Функції, які є частиною функцій керування ресурсами забезпечують відображення інформації, орієнтованої на послуги, яка використовується функціями керування обслуговуванням. Дана інформація буде залежати від ресурсів/технологій, яка використовується ресурсами СПП.

У термінах порівняння зі структурою еТОМ Рекомендацій МСЕ-Т серії М.3050.х, можна відобразити схожість, як показано на Рис.1.8:

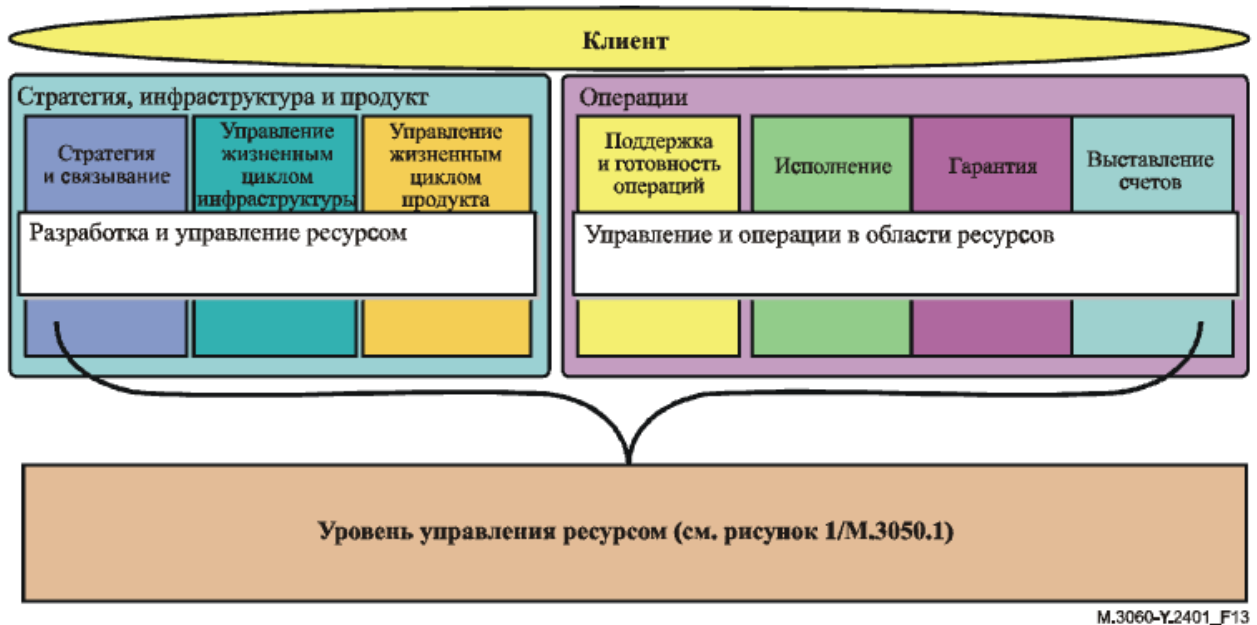


Рис.1.8 13/М.3060/У.2401 – керування ресурсами

Функція керування ресурсами складається з двох головних підфункцій, з'єднаних в місці поділу архітектури СПП на рівень послуг СПП і транспортний рівень СПП:

- функції керування ресурсом обслуговування;
- функції керування транспортним ресурсом.

Функція керування ресурсом обслуговування надає функціональну можливість керування для нового набору характеристик керування ресурсами, що належать до підтримки шару послуг СПП, таких як керування

прикладним програмним компонентом, дані програми, користувачі, дані користувачів, кінцеве обладнання і т. д.

Функція керування транспортним ресурсом надає функціональну можливість для традиційних функцій керування транспортом, з розширенням для забезпечення підтримки транспортного шару СПП, таких як можливість встановлення наскрізного з'єднання IP і керування QoS, і т. д.

Приклад відповідних видів відповідальності функції керування обслуговуваннями і функції керування ресурсами наведено далі. Надання заданого обслуговування кінцевим користувачам приведе до наступних дій:

- створення в функції керування обслуговуванням нового екземпляру послуги, який зв'яже результати розподілу необхідних ресурсів обслуговування і транспортних ресурсів для цього екземпляру послуги за допомогою функції керування ресурсами;

- взаємодій з функцією керування транспортним ресурсом:

- для перевірки наявності необхідних ресурсів транспортної мережі;

- для наскрізної/загальної для всіх прикладних програмних компонентів конфігурації необхідних ресурсів транспортної мережі;

- для конфігурації лінії доступу цього кінцевого користувача відповідно до технічних вимог, відповідними договору про обслуговування;

- взаємодій з функцією керування ресурсом обслуговування:

- для створення всіх даних користувача і відповідної мережевої бази даних в разі нового користувача;

- для створення у відповідній мережевій базі даних всіх відповідних даних обслуговування для цього користувача;

- для розподілу необхідних ресурсів мережі обслуговування;

- для спрацьовування/перевірки конфігурації обладнання в приміщенні користувача (CPE).

Керування ресурсом обслуговування. Функція керування ресурсом обслуговування відповідає за керування ресурсами в шарі послуг СПП.

Функція керування ресурсом обслуговування підрозділяється на функції керування мережею обслуговування і функції керування елементами обслуговування. Інфраструктура шару послуг СПП включає в себе дані, необхідні для забезпечення функціонування послуг СПП з:

- пов'язаними механізмами, використовуваними послугами для доступу до даних;
- керуванням даних.

Функція керування ресурсом обслуговування включає такі функції:

- відображення вимог функції керування обслуговуванням в профілі обслуговування і дані, що інтерпретуються ресурсами нижче;
- керування прикладним програмним забезпеченням і даними прикладного програмного компонента в мережі, включаючи представлення, оновлення, інвентаризацію, поширення, прикладні технології, відкриті інтерфейси програми та пов'язані з ними механізми забезпечення безпеки.

1.3. Мережа LTE формальний опис структури , основні підсистеми та інтерфейси

Мережа оператора мобільного зв'язку – це складна технологічна система, яка складається як з спеціалізованого телекомунікаційного обладнання.

1.3.1. Протоколи доступу LTE E-UTRAN

Швидкість передачі по LTE в низхідному напрямку (до користувача) досягає 100 Мбіт/с, у висхідному - 50 Мбіт/с. Затримка на рівні користувача не перевищує 5 мс за рахунок високої ефективності використання спектра. Такі високі характеристики забезпечуються за рахунок використання

декількох антен (принцип MIMO) і мультиплексування з ортогональним поділом частот OFDM на фізичному рівні.

Мережа E-UTRAN - це найперший вузол у вдосконаленій пакетній системі EPS. Вона забезпечує високу швидкість передачі даних, малу затримку на обох площинах керування і користувача, безшовне перемикання і більше покриття комірки. Розглянемо задачі, функції та процедури рівня доступу в стеку протоколів радіодоступу.

Структура E-UTRAN показана на Рис. 1.9. Мережа складається з вузлів eNodeB (eNB), які забезпечують протоколи площини користувача (PDCP/RLC/MAC/PHY) і керування (RRC). Вузли eNB взаємодіють між собою через інтерфейс X2. Для зв'язку з удосконаленим пакетним ядром (EPC) використовується протокол S1. Обмін з вузлом керування мобільністю (MME) відбувається по інтерфейсу S1-MME, а с обслуговуючим шлюзом (SGW) - по інтерфейсу S1-U. Інтерфейс S1 підтримує зв'язки типу безліч-безліч між MME, SGW і eNB.

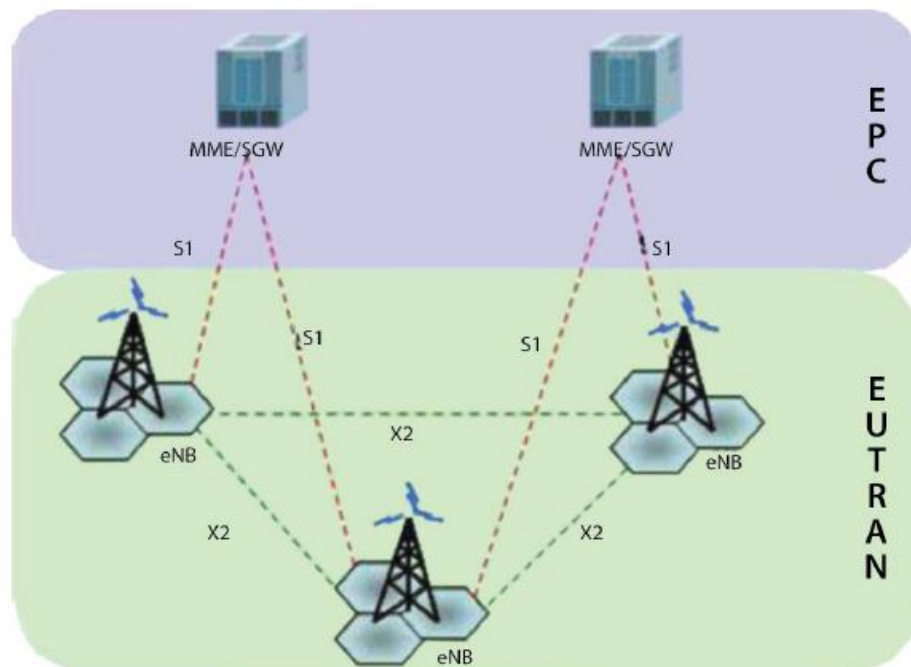


Рис. 1.9 Архітектура мережі E-UTRAN

Інформацію, яку пропускає через себе UTRAN, прийнято розділяти на два шари. До шару доступу (AS) відносяться дані, необхідні для взаємодії

терміналу користувача (UE) і мережі UTRAN. Шар без доступу (NAS) містить інформацію, переноситься між базової мережею оператора (CN) і UE через UTRAN.

Рівень доступу об'єднує протоколи радіодоступу. Це протоколи, що забезпечують спільне використання радіоресурсів обладнання користувача і мережі доступу. Крім того, AS відповідає за з'єднання з каналом радіодоступу (RAB), за допомогою яких забезпечується взаємодія між UE і CN (сервіс NAS).

Рівень доступу надає обладнанню користувача можливість отримання доступу до ресурсів і сервісів мережі, а також всю необхідну інфраструктуру. Протоколи радіодоступу виконують такі функції:

- керування ресурсами радіоканалу (RRM). Це керування радіоканалом і радіоприйомом, контроль мобільності з'єднання і динамічний розподіл ресурсів обладнання користувача в обох напрямках передачі.
- керування трафіком:
- передача даних, в т.ч. в режимі реального часу, між інфраструктурою (рівень NAS) і обладнанням користувача;
- обробка всіх типів даних при різних параметрах каналу (рівень активності, пропускна здатність, затримка передачі і ймовірність появи помилкових бітів);
- ефективне перетворення атрибутів трафіку, які використовують не-LTE прикладними програмними компонентами, в атрибути каналу радіодоступу (RAB) на рівні доступу;
- стиснення IP-заголовка і шифрування потоків даних користувача;
- самостійний вибір MME на обладнанні користувача, коли мережа не надає відповідної інформації;
- передача даних з площини користувача на SGW;
- керування розташуванням: розподіл і передача пошукових повідомлень;
- розподіл і передача широкомовної інформації;

- задання конфігурації вимірюваних параметрів і форми виведення результатів для розподілу ресурсів і забезпечення мобільності;
- розподіл і передача повідомлень про землетруси і цунамі;
- надання первинного доступу до мережі, реєстрація та приєднання до мережі або вихід з неї;
- керування передачею на різних рівнях: між eNodeB, всередині eNodeB, між eNodeB зі зміною MME, між eNodeB зі збереженням MME, але зміною SGW, між RAT;
- функціональна різноманітність і шифрування;
- кодування радіоканалу.

SAE – це архітектура ядра мережі, розроблена консорціумом 3GPP для стандарту бездротового зв'язку LTE (Рис. 1.10).

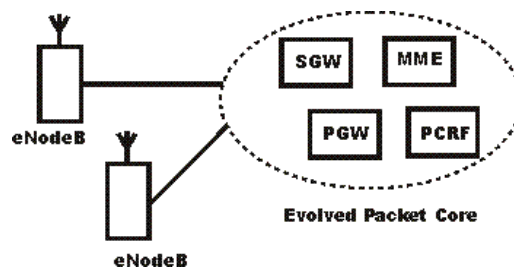


Рис. 1.10 Архітектура ядра мережі SAE

SAE є еволюційним продовженням ядра мережі GPRS, з деякими відмінностями:

- спрощена архітектура - архітектура SAE знижує експлуатаційні та капітальні витрати. Нова плоска модель означає, що потрібно підвищити пропускну здатність вузлів тільки двох типів (базових станцій і шлюзів), щоб вони впоралися з трафіком в разі його значного зростання.
- цілком побудована на IP (AIPN) – перші концепції 3G було розроблено для того, щоб голос, як і раніше передавався по системі з комутацією каналів. З тих пір спостерігався перехід до IP-мереж. Відповідно архітектура SAE побудована на базі IP-мережі.
- забезпечує більшу пропускну здатність мережі радіодоступу (RAN) – передбачається, що низхідний канал (Down Link) буде мати швидкість понад

100 Мбіт/с, і основна увага системи буде зосереджена на мобільності смуги пропускання, мережа повинна буде підтримувати набагато більше рівнів даних.

- забезпечує меншу затримку RAN - зі збільшенням необхідних рівнів взаємодії і більш швидких відповідей, SAE концепція забезпечить рівень затримки в районі 10 мс.

- підтримує мобільність між декількома гетерогенними RAN, що включає підтримку, як систем типу GPRS, так і не-3GPP систем (наприклад WiMAX). [1]

Основним компонентом архітектури SAE є Evolved Packet Core (EPC). EPC служить еквівалентом мережі GPRS. Архітектура EPC показана на Рис. 1.11.

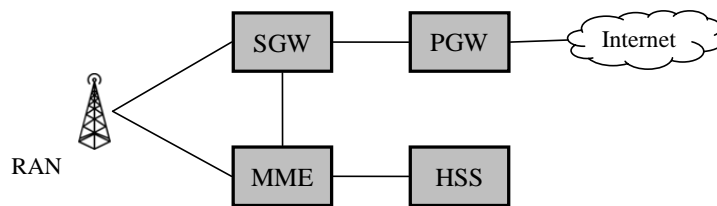


Рис. 1.11 Типова архітектура EPC/LTE

Архітектура EPC є мережею All-IP з комутацією пакетів, який вміщує різні схеми бездротового доступу, такі як Long Term Evolution (LTE). Мережа EPC складається з різних вузлів: Evolved NodeB (eNB) вузла для радіодоступу LTE, об'єкта керування мобільністю (MME) – вузол для керування мобільністю терміналу, вузла домашньої абонентської системи (HSS), вузла бази даних інформації про користувачів, вузла обслуговуючого шлюза (SGW), вузла якірної точки мобільності для керування терміналами, а також вузла шлюза пакетної мережі передачі даних (PGW), який служить шлюзом між терміналами і зовнішніми мережами, такими як Інтернет. [2]

MME – це ключовий контролюючий модуль для мережі доступу LTE. Він відповідає за процедури забезпечення мобільності, хендовера, стеження і пейджінга UE. Він бере участь в процесах активації/деактивації мережевих ресурсів і так само відповідає за вибір SGW для UE при початковому

підключенні і при хендовері всередині LTE зі зміною вузла ядра мережі (CN). Він відповідає за аутентифікацію користувача (при взаємодії з HSS).

SGW – призначений для обробки і маршрутизації пакетних даних, що надходять в підсистему базових станцій. SGW маршрутизує і направляє пакети з одними даними, в той же час виконуючи роль вузла керування мобільністю для даних користувача при хендовері між базовими станціями (eNodeB), а також вузла керування мобільністю між мережею LTE і мережами з іншими технологіями 3GPP. Коли UE вільний і не зайнятий викликом, SGW відключає низхідний канал даних (DL) і створює пейджинг, якщо потрібно передати дані по DL в UE напрямку. Він керує і зберігає стани UE (наприклад вимоги до пропускної здатності для IP-сервісів, внутрішню інформацію мережевої маршрутизації). Він також надає копію даних користувача при узаконеному перехопленні.

PGW – забезпечує з'єднання від UE до зовнішніх пакетних мереж даних, будучи точкою входу і виходу трафіку для UE. UE може мати одночасно з'єднання з більш ніж одним PGW для підключення до декількох мереж. PGW виконує функції захисту, фільтрації пакетів для кожного користувача, підтримку білінгу, узаконеного перехоплення і сортування пакетів. Інша важлива роль PGW - бути вузлом керування мобільністю між 3GPP і не-3GPP технологіями, такими як WiMAX і 3GPP2 (CDMA 1X і EVDO).

HSS – це центральна база даних, яка містить інформацію про користувача. Функції HSS включають мобільність, керування викликами і підтримкою встановлення сеансу, аутентифікацію і авторизацію користувачів. HSS базується на pre-Rel-4 Home Location Register (HLR) і центрі аутентифікації (AuC). [1]

1.3.2. Інтерфейси між вузловими елементами в мережах стандарту LTE

Структура мережі стандарту LTE зазнала значних змін в порівнянні з мережами попередніх поколінь. Це вплинуло також і на зміну інтерфейсів між вузлами мережі. На Рис. 1.12 нижче представлена загальна модель мережі стандарту LTE та її основні інтерфейси.

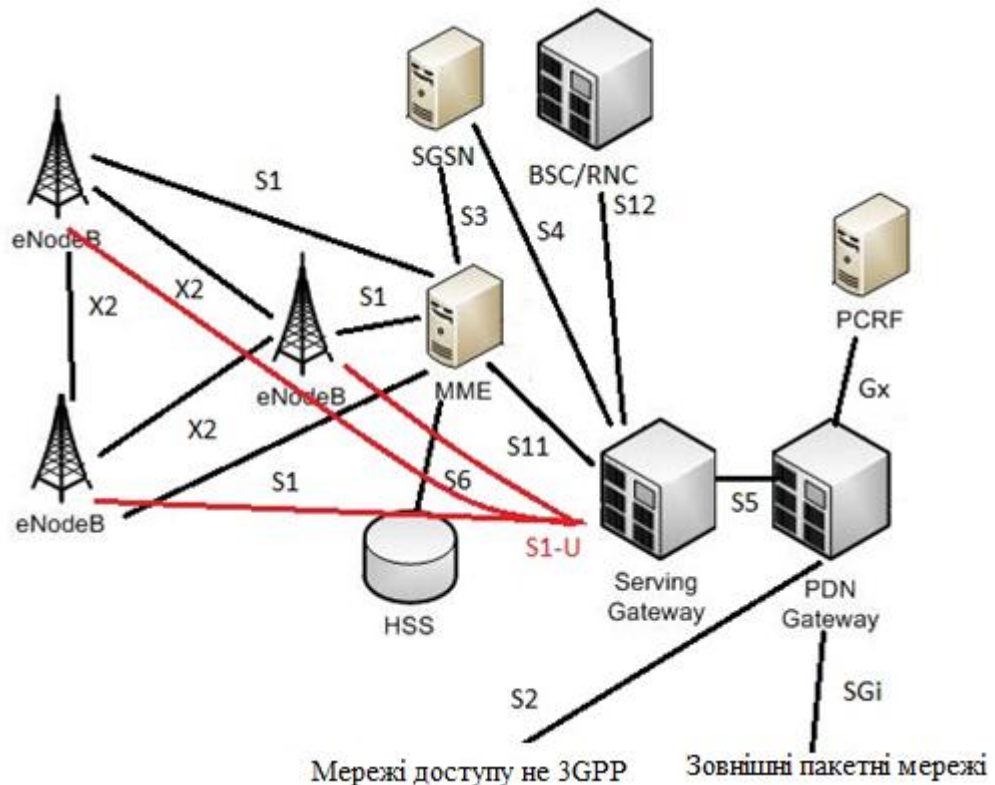


Рис. 1.12 Інтерфейси мережі стандарту LTE

X2- інтерфейс між eNodeB. Базові станції в мережі LTE з'єднані за принципом «кожен з кожним».

S1 - інтерфейс зв'язує підсистему базових станцій E-UTRAN і MME. По даному інтерфейсу передаються дані керування.

S1-U - інтерфейс між E-UTRAN і SAE, по якому передаються дані користувача.

S2 - інтерфейс для організації з'єднання між PDN-Gateway і мережами доступу, які не розроблялись 3GPP.

Адаптація до пакетного терміналу	X	X								
Збір даних					X	X	X	X	X	
Оператор, що визначає заборону					X	X				X
Пакетна маршрутизація і передача										
Relay	X	X	X	X	X	X	X	X	X	
Маршрутизація	X	X	X	X	X	X	X	X	X	
Переклад і зіставлення адрес	X	X		X	X	X	X	X	X	
Інкапсуляція	X	X		X	X	X	X	X	X	
Тунелювання				X	X	X	X	X	X	
Компресія	X	X		X	X					
Шифрування	X	X		X	X					X
Керування мобільністю	X	X			X	X	X	X	X	X
Логічне керування посиланнями										
Створення логічного посилання	X				X					
Обслуговування логічних посилань	X				X					
Керування радіоресурсом	X	X	X	X	X					

1.4. Особливості віртуалізації підсистем мобільного зв'язку LTE

1.4.1. Гетерогенне телекомунікаційне середовище

Гетерогенне телекомунікаційне середовище – організаційно-технічна сукупність, що складається з каналів зв'язку (фізичних та віртуальних), вузлів передачі даних (керованих апаратно та/або програмно) та

обчислювального середовища, розташованого у дата центрах (як власних так і орендованих) з віртуалізованою організацією процесів обробки інформації.

Віртуалізація мережевих функцій (NFV) направлена на перетворення мережевої архітектури операторів за рахунок впровадження стандартної технології віртуалізації для об'єднання багатьох типів мережевого обладнання у високорівневі сервери, комутатори та сховища даних, які можуть бути розташовані в різних точках присутності (NFVI-PoP), таких як дата-центри, мережеві вузли та приміщення кінцевого користувача (ETSI GS NFV 001 V1.1.1(2013-10)).

Загалом, усі мережеві функції та вузли можуть розглядатись для віртуалізації і повинні бути включені за стандартами. Нижче описано основні високорівневі цілі NFV:

- Швидке інноваційне обслуговування через розгортання програмного забезпечення та встановлення чіткого взаємозв'язку між мережевими функціями і кінцевими послугами.
- Покращення операційної ефективності за рахунок спільної автоматизації та робочих процедур.
- Зниження енергоспоживання за рахунок міграції робочих навантажень та вимкнення обладнання, що не використовується.
- Стандартизовані та відкриті інтерфейси між мережевими функціями та їх об'єктами керування, щоб роз'єднані елементи мережі могли забезпечуватись різними гравцями.
- Більша гнучкість у наданні віртуальних мережевих функцій для апаратного забезпечення.
- Підвищення загальної ефективності обслуговування в порівнянні з виділеними апаратними реалізаціями.

1.4.2. Віртуалізація маршрутизатора на границі локальної мережі оператора мобільно зв'язку CPE (vE-CPE) - функціональний опис

Рішення vE-CPE покращує мережу підприємства шляхом заміни пристроїв віртуалізованими рішеннями, сумісними з NFV, розташованими як у корпоративній хмарі, так і в NFV інфраструктурі оператора. Послуги, що надаються vE-CPE, можуть включати в себе маршрутизатор, що забезпечує QoS та інші високотехнологічні послуги, такі як стабільний брандмауер L7, виявлення та попередження вторгнень тощо. Прискорювачі програм також розгортаються або як окремі пристрої, або як інтегровані сервіси маршрутизатора.

На Рис. 1.13 представлено типове телекомунікаційне підприємство, що включає штаб-квартиру з централізованою корпоративною IT-інфраструктурою та декілька філій, з'єднаних один з одним та з головним офісом підприємства. Функціональність VE-CPE може розташовуватись в різних місцях.

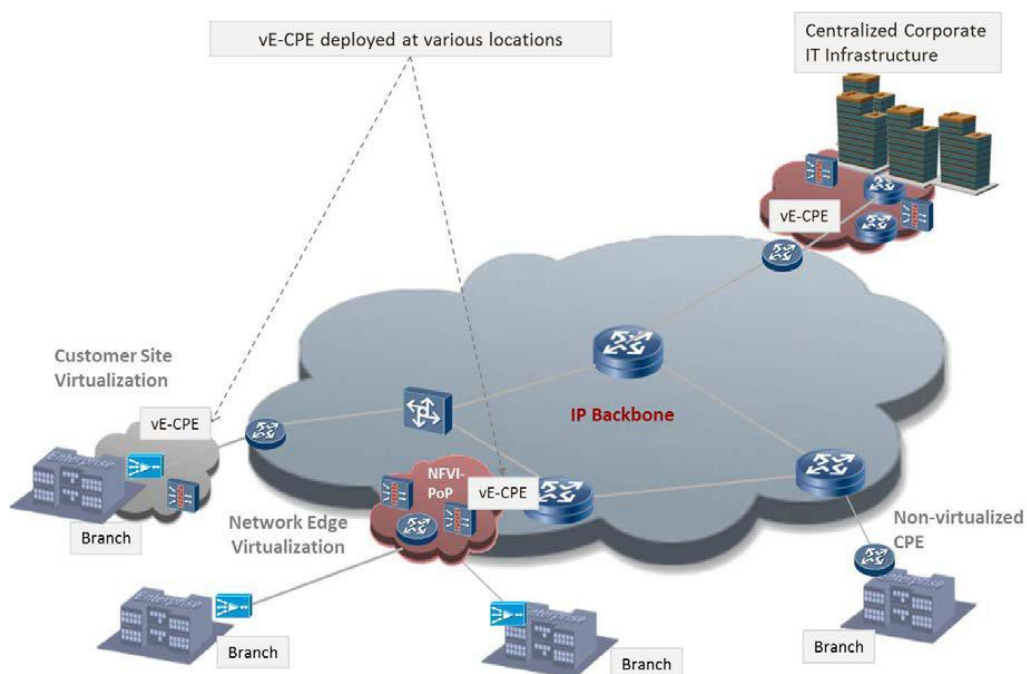


Рис. 1.13 Приклади розташування vE-CPE

На Рис. 1.13 представлено перерозподіл функціональних можливостей в результаті віртуалізації CPE. Локальний трафік підприємства обробляється

локальним маршрутизатором L2 або L3, що забезпечує фізичний зв'язок (і, можливо, додаткову функціональність), а локальна мережа підприємства поширюється на мережу NFV оператора, розташовану у vE-CPE. Приклад функціональності, яку надає vE-CPE, включає в себе маршрутизацію, припинення VPN, підтримку QoS, DPI, NG-FW та WOC (WAN Optimizer Controller). На Рис. 1.14 порівнюється невіртуальний клієнтський сайт, який обслуговує не віртуалізований CPE, і веб-сайт, який обслуговує vE-CPE. Пунктуальні фіолетові лінії вказують, де може розташовуватись функціонал vE-CPE.

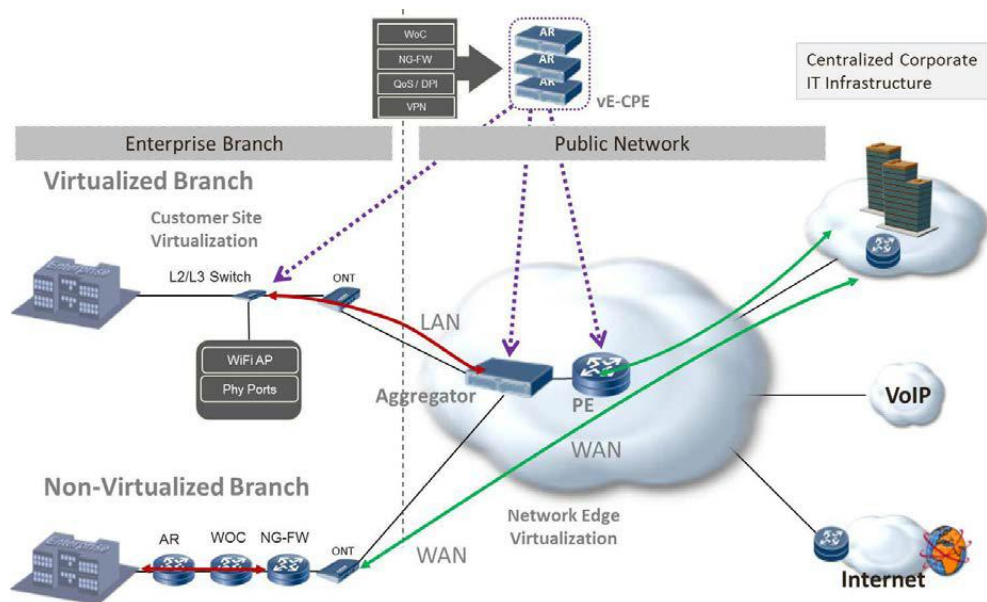


Рис. 1.14 Не віртуалізоване CPE та vE-CPE

1.4.3. Віртуалізація порогового маршрутизатора оператора (vPE) - функціональний опис

Віртуалізація базових маршрутизаторів може бути неможливою у короткостроковій перспективі через високі вимоги до пропускної здатності, але віртуалізація порогового маршрутизатора PE з додатковими перевагами забезпечення масштабованості послуг, наданих постачальником є більш імовірною, за рахунок динамічної зміни об'єму або розподілу віртуальних ресурсів. Провайдер, який надає послуги віртуальної приватної мережі

(PPVPNS) у vPE включає IP-VPN 3-го рівня, VPLS 2-го шару, EVPN тощо [i.8].

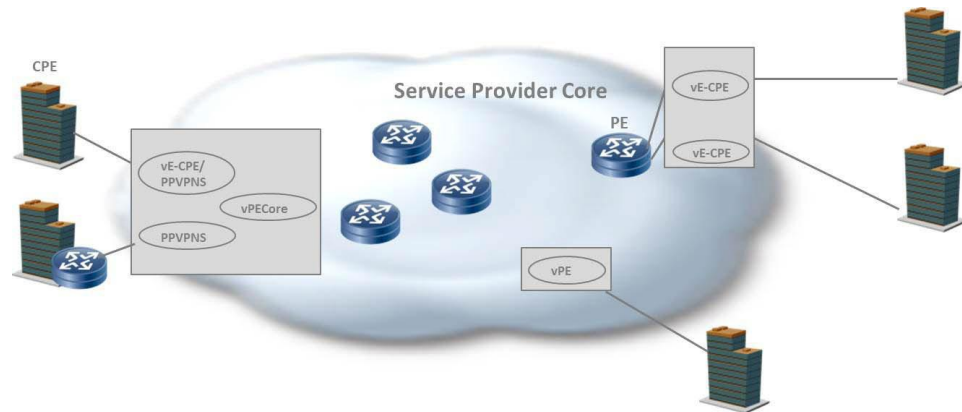


Рис. 1.15 Віртуалізація CPE, віртуальних мережесервісів та основних функцій порогового маршрутизатора

Віртуальні функції PE можуть:

- Інтегруватись в єдину віртуальну машину (vPE на Рис. 1.15), реалізувати еквівалентну функціональність єдиного фізичного PE або підмножину фізичних PE для масштабованості та продуктивності.
- Розбиватись на основний набір функцій (vPEcore на Рис. 1.15) та функції віртуальних мережесервісів, які включають або не включають функції CPE (PPVPNS і vE-CPE / PPVPNS).

Існує багато мережесервісів, які зазвичай сьогодні розгортаються в корпоративних мережах в якості спеціалізованої інфраструктури обладнання, де в майбутньому постачальник послуг може надавати VNFaaS для підприємства. Ці мережесервіси Enterprise включають в себе:

- 1) AR - маршрутизатор доступу для підприємств / Enterprise CPE
- 2) PE - пороговий маршрутизатор
- 3) FW - корпоративний брандмауер
- 4) NG-FW – брандмауер наступного покоління
- 5) WOC - контролер оптимізації мережі WAN корпоративного рівня
- 6) DPI - Глибока перевірка пакетів (пристрій або функція)
- 7) IPS - Система запобігання вторгненню та інші пристрої безпеки

8) Моніторинг продуктивності мережі.

За оцінками, велика кількість віртуалізованих пристроїв потребує підтримки з боку порогового маршрутизатора, що вимагає величезної кількості ресурсів від NFVI. VPE повинен мати можливість самостійного масштабування на площині даних та площині керування, щоб підтримувати великі таблиці переадресацій та велику кількість потоків. У порівнянні з рішеннями Virtualized Home, для віртуалізації підприємства потрібна значна менша кількість VNF, кожна з яких має набагато більшу кількість потоків та вимог до продуктивності.

Щоб обмежити вартість та масштаб, велика кількість віртуалізованих пристроїв повинна бути інтегрована на обмежену кількість процесорів.

Для досягнення цілей ефективності віртуальні функції порогового маршрутизатора повинні:

- Інтегруватись в єдину віртуальну машину;
- Розбиватись на основний набір функцій та віртуальних мережевих послуг.

VPE повинно мати можливість динамічного масштабування, щоб підтримувати велику таблицю переадресації і велику кількість потоків. Досягти масштабування vPE можна шляхом:

- Модифікації ресурсів інфраструктури, що виділяються на екземпляр vPE (наприклад, збільшення пам'яті);
- Створення додаткових екземплярів vPE.

У віртуалізованому середовищі відповідальність за належне виконання кожного сценарію лягає на постачальника VNFaaS. VE-CPE і vPE необхідні для підтримки великої кількості прикладних програмних компонентів і послуг, керованих динамікою підприємства. Крім того, під час переходу від звичайних до віртуалізованих мереж буде багато топологій та конфігурацій мереж.

Як постачальник VNaaS, так і користувач поділяють відповідальність за керування vPE та vE-CPE. Користувачі підприємств будуть керувати та налаштовувати свої пристрої CPE і керувати версіями SW після модернізації, навіть якщо вони є віртуалізованими та надаються в якості служби.

VNaaS вводить єдину точку збою, оскільки внутрішні операції з підприємством, які залежать від CPE, можуть не працювати належним чином після втрати мережових з'єднань. Завдання полягає в тому, щоб забезпечити безперервність роботи на підприємстві під час відмови мережі або доступу до мережі (відповідно до поточної поведінки мережі).

Якщо vPE і vE-CPE повинні контролюватися централізованим контролером, слідуючи принципам і стандартам архітектури SDN, то надійне підключення між контролером та віртуальними пристроями, незалежно від їх розташування, має вирішальне значення.

Віртуалізоване середовище має гарантувати повну ізоляцію серед користувачів. Розширення корпоративної локальної мережі в операційну мережу вимагає встановлення з'єднання VPN між підприємством та віртуальною функцією оператора. Особливі заходи необхідні для захисту корпоративних даних та файлів конфігурації.

Надання VNaaS в якості служби оцінки вимагає вимірювання показників та інфраструктури, що відповідає типу VNF, а також відповідних угод про рівень обслуговування. Значення показників якості обслуговування, отриманні за допомогою служби VNaaS потребуватимуть відповідного аудиторського методу обліку, який буде використовуватися в якості основи для платіжних послуг.

1.4.4. Граф переадресації VNF

Граф переадресації мережевої функції (NF) [1] визначає послідовність мережових функцій, за якою перетинаються пакети. Проста мережева служба

[1] може бути реалізована в середовищі NFV за допомогою зв'язків point-to-point. Цей сценарій використання показує, що можуть знадобитись більш складні структури, такі як граф переадресації VNF (VNF FG) [1].

Графи VNF є аналогами з'єднань існуючих фізичних пристроїв кабелями. Кабелі є двонаправленими, тому більшість технологій мережевих даних (наприклад, Ethernet) будуть найближчим часом використовуватись у віртуалізованих розгортаннях. Іншими словами, VNF Forwarding Graph забезпечує логічне підключення між віртуальними пристроями (наприклад, VNFs).

Щоб реалізувати цілі NFV, постачальникам послуг необхідно розвинути мережеві служби на абстрактному рівні, а потім розгорнути їх в екземплярах, пов'язаних з певними ресурсами NFVI (обчислювальні вузли, точки завершення мережі інфраструктури, існуючі фізичні NE, і т. д.).

Віртуальні мережеві функції у VNF FG мають стандартизовані інтерфейси (наприклад, L1, L2, L3, L4 та / або L7). У деяких із них, пакети мають певне призначення (наприклад, серверні функції), а в інших – ні (наприклад, Інтернет). Багато інших сценаріїв використання мають ті ж характеристики, що і сценарій використання VNF FG, а вимоги, архітектура та специфікації цих загальних характеристик повинні відповідати цілям NFV для забезпечення переходу від існуючих фізичних мережевих функцій до віртуальних аналогів.

Актори та ролі: VNF Provider - це постачальник програмного забезпечення для керування VNF та NFV; оркестрування - набір операційних систем, що підтримують NFVI.

Сценарій використання графа переадресації VNF має наступні логічні частини та відносини між суб'єктами (Рис. 1.16):

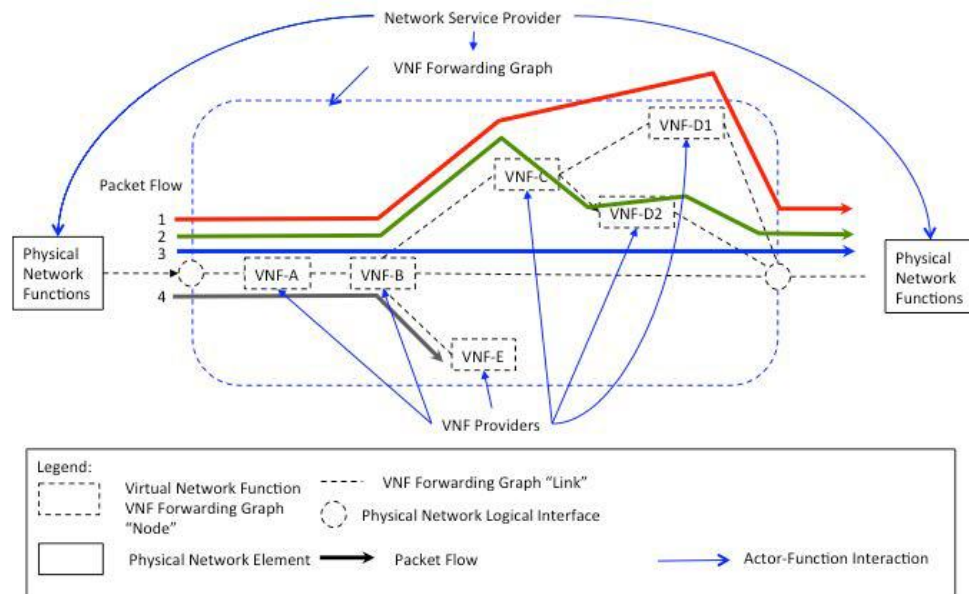


Рис. 1.16 Логіка графа переадресації VNF

Функція фізичної мережі: реалізація, яка є частиною загальної невіртуалізованої служби, що розгортається та керується постачальником послуг мережі. Це може бути фізичний доступ або магістральна мережа, автономна VM, точка з'єднання між кількома FGF VNF, що надаються різними доменами адміністратора (наприклад, NSPs).

Фізичний мережевий логічний інтерфейс: межа між VNF FG та фізичними мережевими функціями визначається постачальником послуг. Він може ґрунтуватися на полях у заголовку пакетів, які є джерелом або призначенням надходження пакетів від функції фізичної мережі.

Пакетний потік: вихід мережі, який є частиною загальної послуги і полягає в тому, що певні групи пакетів ідуть по тому ж шляху через VNF FG. Функціональність VNF, конфігурація та стан визначають потік пакетів через граф переадресації VNF, а пропущені VNF можуть відрізнятися в кожному напрямку для пакетів одного і того ж двонаправленого потоку.

Інфраструктура мережі NFV: забезпечує послуги з'єднання між VNF, які реалізують зв'язки графів пересилання між вузлами VNF в апаратному та/або програмному забезпеченні, як показано червоними стрілками. Вона може містити функції, включаючи класифікацію трафіку, інкапсуляцію тунелю, керування рухом та/або деякі форми балансування навантаження.

На Рис. 1.16 наведено приклад графа переадресації VNF, який постачальник послуг може використовувати в якості частини свого дизайну обслуговування.

У цьому прикладі постачальник послуг розробив кінцеву мережеву службу між двома функціями фізичної мережі, що включає кілька віртуальних мережевих функцій (VNF-A, VNF-B, VNF-D1, VNF-D2, VNF-E). Ці VNF були надані одним або більше провайдерів VNF. У цих VNF є деякі метадані, пов'язані з ними, які описують основні характеристики VNF. Фактична мережева послуга являє собою набір усіх можливих пакетних потоків, які перетинають VNF FG та будь-які PNF, наприклад, як показано на Рис. 1.16. Служба мережі включає в себе інформацію (а також логіку у самих VNF), які використовують VNF FG.

Логічний сценарій використання VNF FG пов'язаний з фізичними елементами та додатковими відносинами між собою, як показано на прикладі Рис. 1.16. В зв'язку з цим вводиться наступна термінологія:

Фізичний зв'язок мережі: відносини між мережевою інфраструктурою NFV та фізичним мережевим портом на мережевої функції, на кордонах між VNF та фізичними елементами.

Фізичний мережевий порт: фізичний порт на фізичній мережі або фізичний мережний комутатор/маршрутизатор або фізичний NIC.

Мережевий шлях переадресації: послідовність портів комутації обладнання та програмного забезпечення і операцій в інфраструктурі мережі NFV, який налаштовується керуванням та оркеструванням, що реалізує логічні інтерфейси "вузла" і зв'язує вузол VNF логічного графа переадресації VNF, (наприклад, VNIC на VM) . Інформація VNF FG описує характеристики цих "посилань". Традиційні способи впровадження графів переадресації в мережі включають в себе: пересилання між фізичними пристроями на основі фізичного інтерфейсу, мережеві мости на базі VLAN, підмережі IP, конфігурації тунелів, маршрутизацію на основі політики та конкретні конфігурації BGP. Перемикач керування SDN (наприклад, OpenFlow) може

реалізовувати ці традиційні методи, але також може безпосередньо створювати граfi переадресації мережі різними динамічними способами.

Середовище віртуальних машин: характеристики обчислювального та мережевого середовища для певного набору елементів програмного забезпечення VNF, налаштованих за допомогою керування та оркестрування. Воно визначається інформацією, наданою постачальником VNF, та інформацією, наданою постачальником послуг мережі для VNF FG.

Постачальник послуг повинен мати можливість інвентувати всі ці VNF у своїх NFVI і прогнозувати діапазон очікуваної поведінки та продуктивності кінцевої служби мережі. На Рис. 1.17 показаний мережевий сервіс, що складається з VNF між двома фізичними мережевими функціями, при яких трафік пересилається через два фізичні пристрої та два VNF-файли (VNF-A, VNF-B). У цьому прикладі VNF-A є цілком віртуалізованою мережевою функцією, так як мережеве підключення також віртуалізується комутатором, однак VNF-B лише частково віртуалізується з трафіком площини даних, що проходить через фізичний перемикач.

VNF FG певного типу, де вузли та посилання мають подібну топологію з атрибутами, що визначаються параметром (наприклад, ємність, обмеження продуктивності), повинні використовувати загальний шаблон.

Надання графа VNF означає, що конкретний об'єкт VNF FG відповідно до цього шаблону повинен представлений в якості екземпляру графа переадресації NFV для набору потоків (наприклад, споживачі, підприємства, користувачі бездротового доступу, що мають доступ до Gi LAN тощо).

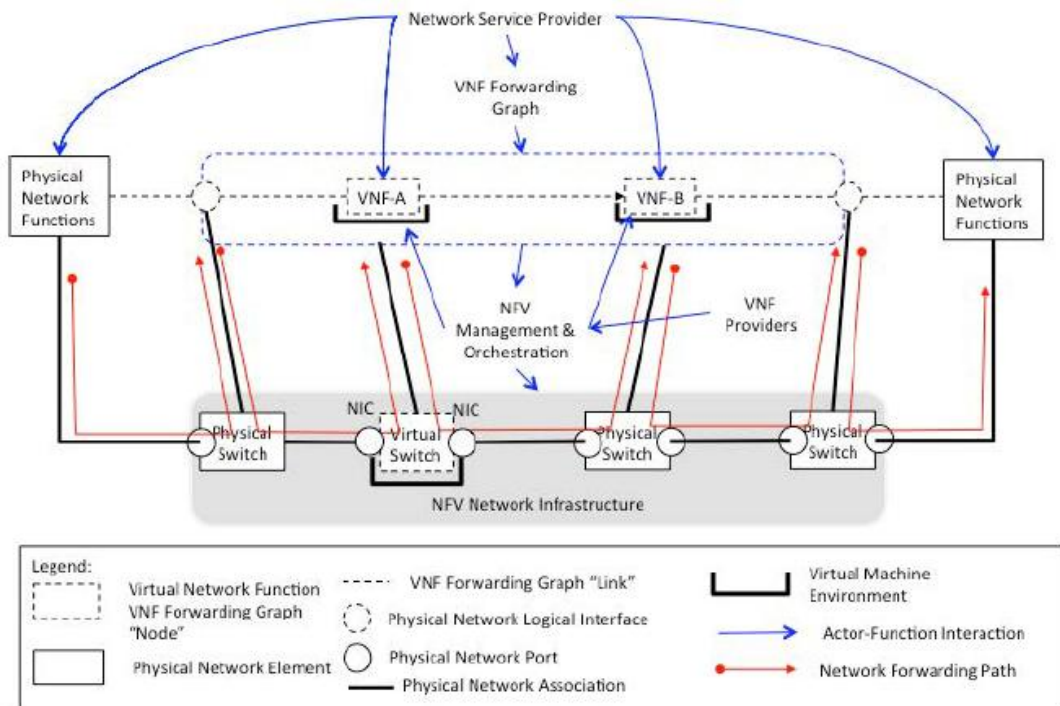


Рис. 1.17 Фізична архітектура VNF FG

Коли забезпечується мережева служба, NFV Framework повинна вести облік використаних ресурсів інфраструктури, щоб майбутні операційні процеси (такі як локалізація несправності, відновлення, зміна розміру або припинення дії послуги) могли бути здійснені на всіх відповідних об'єктах в VNF FG.

1.4.5. Віртуалізація базової мобільної мережі та IMS

Мобільні мережі завантажені широким спектром власних пристроїв. Основна задача віртуалізації мережевих функцій полягає у зменшенні складності мережі та пов'язаних з цим проблем за рахунок стандартних технологій віртуалізації для консолідації різних типів мережевого обладнання. Така консолідація обладнання повинна зменшити сукупну вартість володінь (TCO – Total Cost of Ownership). Гнучкий розподіл мережевих функцій на такий апаратний пул ресурсів може значно підвищити ефективність роботи мережі. Це також дозволяє реагувати на збільшення попиту на певні послуги (наприклад, голосовий зв'язок), не покладаючись на

механізми контролю обмеження викликів у великомасштабному сценарії стихійних лих, таких як землетрус Великого Східного Японії, під час якого мобільні мережі зіткнулися з величезною кількістю спроб дзвінків. Можливими перевагами віртуалізації мобільних базових мереж та IMS є:

- Зниження показнику TCO.
- Підвищення ефективності використання мережі завдяки гнучкому розподілу різних мережевих функцій на апаратний ресурсний пул.
- Підвищення якості обслуговування та надійності для кінцевих користувачів шляхом динамічної реконфігурації мережі, властивій технологіям віртуалізації.
- Еластичність: ємність мережі, призначена для кожної мережевої функції, може бути динамічно модифікована відповідно до фактичного навантаження в мережі, тим самим збільшуючи її масштабованість.
- Реконфігурація топології: топологію мережі можна динамічно переконфігурувати для оптимізації продуктивності.

3GPP – стандартна організація, яка визначає мережеву архітектуру та специфікації для мережевих функцій (NF) мобільних і конвергентних мереж.

В покращеному пакетному ядрі (Evolved Packet Core – EPC), що являє собою новітню мережеву архітектуру стільникової системи, прикладами мережевих функцій є MME, S/P-GW та ін.

В мультимедійній підсистемі IP (IMS), яка є архітектурою керування сеансами для підтримки надання мультимедійних послуг через EPC та інші IP-мережі, прикладами мережевих функцій є P-CSCF, S-CSCF та ін. HSS та PCRF також є мережевими функціями мережі 3GPP™, необхідними для роботи в архітектурі end-to-end EPC та IMS для надання послуг. Аналогічно, онлайн-і оффлайн- системи тарифікації (OCS та OFCS) - це системи, які фіксують записи про тарифікацію як частину керування сеансами.

Цей випадок має на меті застосування віртуалізації до EPC, IMS та інших мережевих функцій, згаданих вище.

На Рис. 1.18 представлено можливий вигляд віртуалізації EPC на основі NFV.

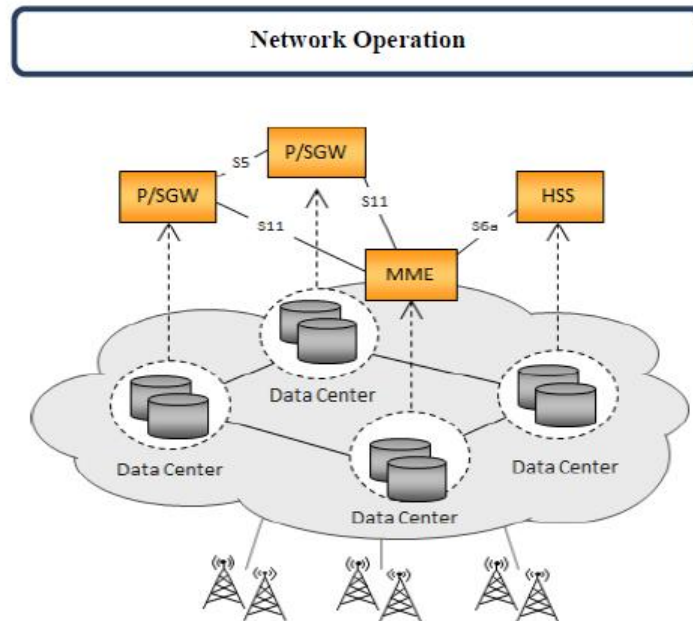


Рис. 1.18 Віртуалізація EPC

Віртуалізовані мережеві функції (VNF), наприклад, S/P-GW, MME можуть масштабуватися незалежно залежно від їх вимог до ресурсів. Наприклад може виникнути ситуація, коли необхідно збільшити ресурси платформи користувача, не впливаючи при цьому на площину керування, і навпаки. Крім того, VNF, що займаються платформою даних, можуть вимагати іншої кількості ресурсів NFV Infrastructure, ніж ті, що працюють тільки з сигналізацією.

Різні сценарії можуть бути увімкнені, якщо, наприклад, все ядро EPC є віртуалізованим в одному NFVI-PoP або лише деякі мережеві функції є віртуалізованими.

Для того, щоб досягти необхідної безперебійної роботи служби оператора та наявності сервісу, необхідно забезпечити стійкість як на площині керування, так і на площині даних. Оскільки віртуалізація дозволяє розв'язувати NF з основного апаратного забезпечення, створення нових схем пружності стає можливим завдяки використанню портативних екземплярів VNF, але не обмежуючись переміщенням VM, реплікацією тощо. На Рис.

1.19 показані віртуалізовані EPC та IMS, де функції S/P-GW та IMS працюють з підключенням PDN та сеансом IMS, відповідно. Коли динамічне переміщення цих екземплярів VNF здійснюється через перевантаження або несправності віртуальної машини автоматично або за вимогою, переміщення керованих сеансів та/або підключень повинні оброблятися належним чином, щоб забезпечити безперервність роботи оператора і доступність сервісу.

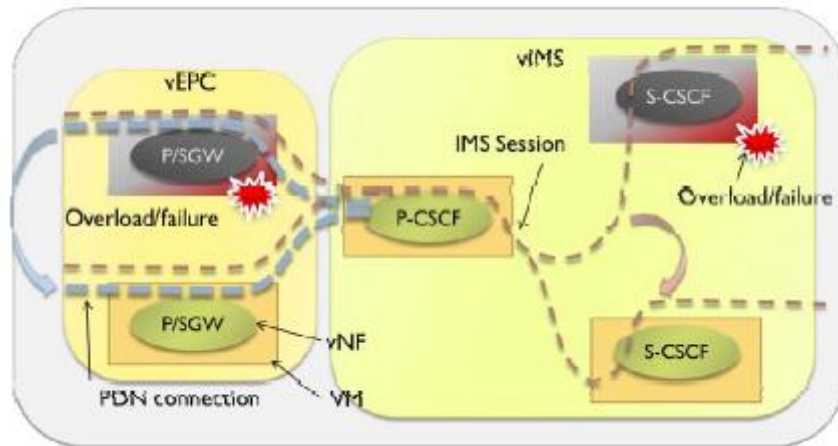


Рис. 1.19 Переміщення VNF

Міжоператорський зв'язок та графік пересилання VNF – потенційні елементи для подальшого вивчення в даному сценарії використання.

Віртуалізована базова мобільна мережа на основі NFV співіснує з невіртуалізованою (Рис. 1.20), оскільки вже розгорнуті мобільні базові мережі не функціонують на NFV. Мережеві оператори повинні мати право самостійно вибирати розгортання NFV відповідно до бажаного плану міграції з невіртуальної до віртуалізованої мережі на основі NFV.

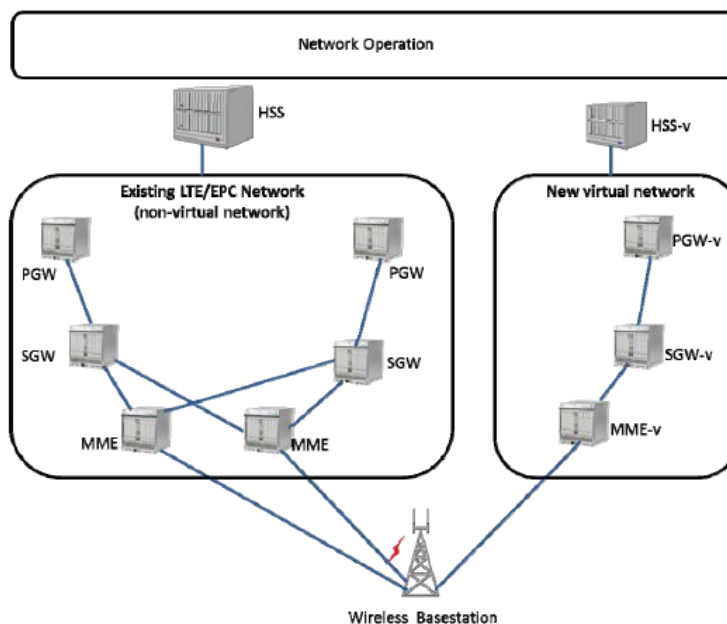


Рис. 1.20 Приклад співіснування віртуальної на невіртуальній базовій мобільній мережі

Залежно від вибору операторів можуть бути різні сценарії взаємодії. В якості прикладів, нижче наведено два сценарії:

- Віртуалізація деяких компонентів мобільної базової мережі. У цьому випадку віртуалізовані лише деякі мережеві функції (Рис. 1.21). Це можуть бути функції керування EPC (наприклад, MME / SGSN), HSS або сервісними вузлами (наприклад, IMS).

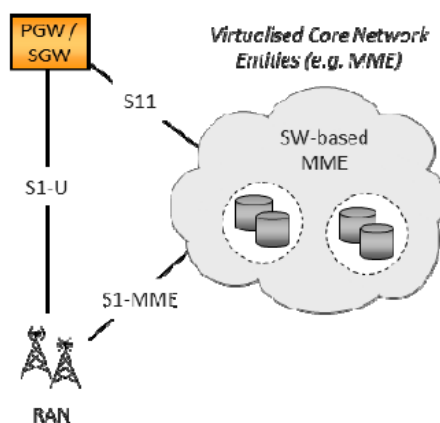


Рис. 1.21 Часткова віртуалізація базової мобільної мережі

- Співіснування віртуалізованої та невіртуальної мобільної базової мережі.

У цьому випадку оператор розгортає повністю віртуалізовану базову мережу, маючи при цьому невіртуалізовану (Рис. 1.22). Віртуалізоване ядро може використовуватися для певних служб та/або пристроїв або для трафіку, що перевищує місткість невіртуалізованої мережі.

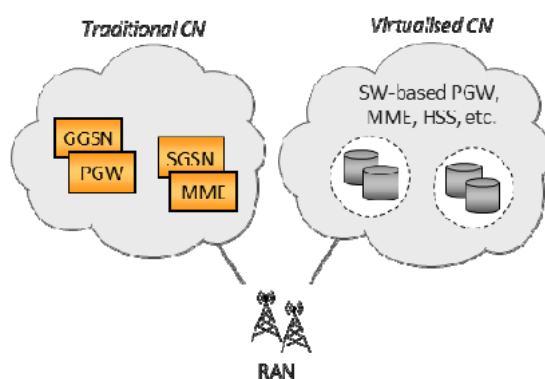


Рис. 1.22 Сервісна віртуалізація базової мережі

Для сценаріїв, що передбачають співіснування віртуалізованих та невіртуальних мобільних базових мереж потрібно визначити наступні політики:

- 1) Мережа радіодоступу (RAN): де збігається віртуальне і невіртуальне мобільне ядро.
- 2) Мережеві операційні системи: яким чином мережева операційна система для невіртуальної мережі взаємодіє з конкретними операціями базової мобільної мережі та чи потрібні нові системи підтримки операцій або вдосконалення існуючих систем.
- 3) Повернення до невіртуалізованої мережі: коли необхідно застосовувати механізм відмовостійкості для невіртуалізованих мережевих функцій.

Віртуалізація мобільної базової станції

Мобільний трафік мережі значно збільшується за рахунок попиту, який генерується застосуванням мобільних пристроїв, в той час як середню кількість прибутку за одного користувача (ARPU) збільшити складно. Тому технологія 3GPP LTE направлена на задоволення потребам в підвищенні швидкості передачі даних та якості обслуговування з невисокою складністю

та незмінним зниженням вартості радіодоступу і пакетного ядра. LTE також розглядається як частина доступу до радіосистем EPS (Evolved Packet System), необхідна для виконання вимог високої ефективності спектра, високої пікової швидкості передачі даних та гнучкості частоти в мережі радіодоступу (RAN). Щоб зберегти прибуток, оператори мобільного зв'язку повинні зменшувати CAPEX/OPEX, а також постійно розвивати та надавати кращі послуги своїм клієнтам. Більша частина витрат електроенергії припадає на вузли RAN. Велика кількість вузлів RAN, таких як eNodeB, зазвичай базується на власних платформах і їх недоліками є тривалий життєвий цикл розробки, розгортання та експлуатації.

Віртуалізація мобільних базових станцій використовує технологію віртуалізації ІТ для реалізації принаймні частини вузлів RAN на стандартних ІТ-серверах, сховищах та комутаторах. Це забезпечить зменшення відбитку та споживання енергії, пов'язані з динамічним розподілом ресурсів та балансуванням навантаження трафіку.

Крім того, NFV дозволяє створювати конкурентне середовище для постачання інноваційних сторонніх мережевих прикладних програмних компонентів, розблокуючи власні межі вузлів мобільних базових станцій.

В основних мережах мобільних операторів, кілька вузлів RAN від різних постачальників, як правило, експлуатуються з різними системами мобільної мережі, наприклад 3G, LTE і WiMAX, які знаходяться в тій же зоні. Передбачається, що ці численні платформи будуть об'єднані у фізичну базову станцію (БС) на основі технологій віртуалізації.

Віртуалізація базової станції (БС) може забезпечувати обмін ресурсами між декількома логічними вузлами RAN з різних систем, динамічно розподіляючи ресурс, а також зменшуючи споживання енергії. Централізована технологія RAN (C-RAN) з віртуалізацією може сприяти більш ефективному використанню ресурсів різних фізичних БС (Рис. 1.23).

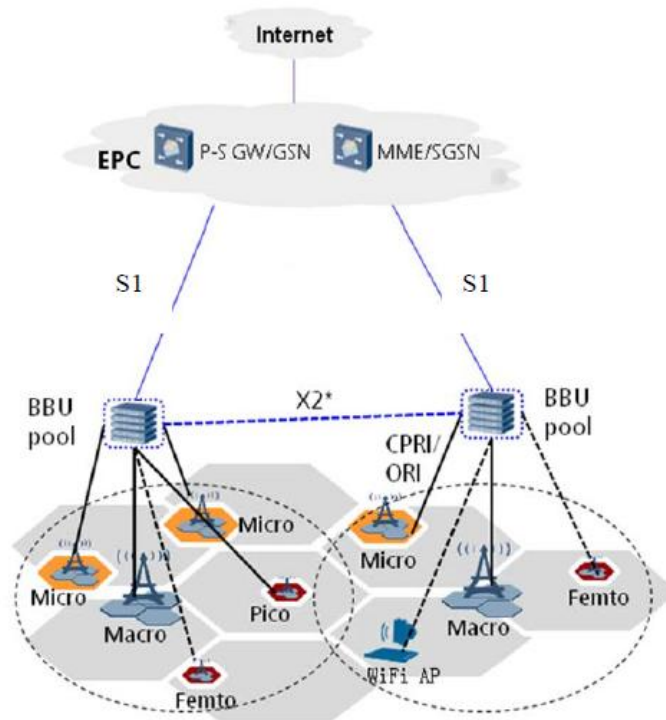


Рис. 1.23 Еволюція архітектури LTE RAN централізованим пулом BBU

БС (використовується тут як загальний термін для позначення базових станцій 2G, 3G Node B і 4G eNode B) є частиною еталонної моделі 3GPP і включає в себе функції радіодоступу. У LTE базові станції відповідають за функції PHY, MAC, RLC (радіозв'язок), RRC та PDCP (протокол конвертації пакетних даних). Рівень PHY містить найбільш обчислювальні інтенсивні завдання, такі як кодування/декодування каналів, FFT / iFFT.

Віртуалізація БС вимагає обробки базової радіопередачі за допомогою технологій візуалізації, таких як високопродуктивні процесори загального призначення та віртуалізація обробки в режимі реального часу для забезпечення необхідної потужності обробки сигналу.

Крім того, віртуалізація БС для C-RAN потребує створення ресурсу обробки, тобто пулу обробки базових частот (BBU) для об'єднання ресурсів в централізоване віртуалізоване середовище, наприклад, дата центр або хмарну інфраструктуру.

Віртуалізація мобільної базової станції повинна підтримувати часткові сценарії розгортання, які враховують різні функції та елементи в частині

RAN різних систем мобільної мережі. Серед можливих випадків використання виділяють:

1) Віртуалізована традиційна базова станція eNodeB та невіртуалізована eNodeB:

Віртуалізована і невіртуалізована eNodeB зв'язуються один з одним стандартним інтерфейсом X2, і мало ймовірно, що існують проблеми інтероперабельності, якщо вони відповідають вимогам 3GPP затримки та джиттера.

2) Віртуалізований пул BBU та невіртуалізована eNodeB:

Віртуалізований BBU повинен мати стандартний інтерфейс X2 з невіртуальним BBU, навіть якщо інтерфейс X2 замінено власним для досягнення більш ефективного спілкування в басейні BBU. Невіртуалізована eNodeB - це фізична базова станція, географічно відокремлена від басейну BBU. У цьому випадку вищезазначені проблеми продуктивності можуть також статися через віртуалізацію та топологічну ситуацію.

Всередині вузла RAN можуть існувати цільові апаратні засоби, оскільки всі функції обробки основної смуги не можуть бути ефективно реалізовані на програмному забезпеченні. Між прискорювачем та стандартною ІТ-платформою повинен бути реалізований інтерфейс API.

1.4.6. Віртуальна мережева функція як сервіс

Телекомунікаційні сервіси (службові) – це послуги, які надаються у гетерогенному телекомунікаційному середовищі *в процесі організації взаємодії* між абонентами або машинами.

Сьогоднішні підприємства розгортають у своїх філіалах по декілька сервісів одночасно. Для багатьох із них підтримка виділеного автономного пристрою є надто затратною та негнучкою. В інших реалізаціях, функціональність може бути забезпечена інтегрованим маршрутизатором доступу, який може бути обмеженим у наборі функцій. По мірі того, як

підприємство розвивається, все більше послуг та прикладних програмних компонентів мігрують до корпорації дата-центрів або загальнодоступних хмар, що приводить до змін в побудові корпоративних мереж. Крім того, мобільність і принцип BYOD (Bring Your Own Device – принеси свій власний пристрій) стають все більш затребуваними, в результаті чого стають актуальними послуги запобігання витоку даних.

Зіткнувшись із необхідністю великих інвестицій, багато підприємств починають шукати альтернативні варіанти. Ці альтернативи можуть включати в себе віртуалізацію Enterprise CPE (маршрутизатора доступу) в мережу оператора.

Такі тенденції віртуалізації в поєднанні з перевагами які надає NFV забезпечують значні бізнес-можливості для постачальників послуг, які намагаються відповідати зростаючим потребам клієнтів. Традиційні IP-маршрутизатори, засновані на власному апаратному та програмному забезпеченні, є одними з найбільш капіталомістких частин інфраструктури постачальників послуг. Маршрутизатори оператора (Provider Edge routers) вичерпують ресурси платформи керування, перш ніж закінчуються ресурси платформи даних, таким чином віртуалізація функцій платформи керування покращує масштабованість.

Зберегти ресурси можна також переміщаючи функціональність маршрутизації від цільових маршрутизаторів до еквівалентних функцій, реалізованих в апаратних середовищах COTS, що забезпечують можливості хмарних обчислень, такі як NFVI.

Замість того, щоб інвестувати власний капітал у розгортання мережевої інфраструктури, постачальник послуг може надавати розширені мережеві функції. Постачальник може використовувати об'єкт VNF за допомогою інтерфейсу NFVI, який забезпечує функціональність, необхідну для впровадження клієнтського обладнання CPE та інший об'єктів VNF для платформи керування маршрутизатором оператора, покращуючи його масштабованість. Створення функціональності VNF для підприємства в

якості служби можна порівняти з поняттям хмарного обчислення – «Програмне забезпечення як сервіс».

Інститут стандартизації NIST SP 800-146 [i.3] визначає програмне забезпечення як сервіс (SaaS) як можливість споживачів використовувати прикладні програмні компоненти, що працюють у хмарній інфраструктурі. Споживач може керувати програмою лише з точки зору конфігурації та не може контролювати базову інфраструктуру.

У цьому прикладі віртуалізованих корпоративних служб VNF є прикладним програмним компонентом постачальника послуг. Підприємство є споживачем послуги. Підприємство не керує та не контролює NFVI або VNF. Підприємство як споживач VNFaaS не має інвестувати додатковий капітал у розширені мережеві функції, надані за допомогою платформи керування, а може отримати їх за певні кошти від постачальника послуг, якщо це необхідно. Постачальник послуг може масштабувати ресурси NFVI, виділені на екземпляр VNF у відповідь на збільшення використання VNF.

NIST SP 800-146 [i.3] визначили наступні переваги моделі SaaS, які також повинні застосовуватися у випадку з VNFaaS:

- Незначна роль програмного забезпечення підприємства для доступу до сервісу;
- Ефективне використання ліцензій на програмне забезпечення;
- Централізоване керування;
- Економія на попередніх витратах

Мережі постачальників послуг Pre-NFV містять пороговий маршрутизатор PE на межі ядра перед пристроєм клієнтського обладнання (CPE), як показано на Рис. 1.24. При цьому є дві бізнес-моделі; як постачальник послуг так і підприємство можуть володіти та керувати CPE.

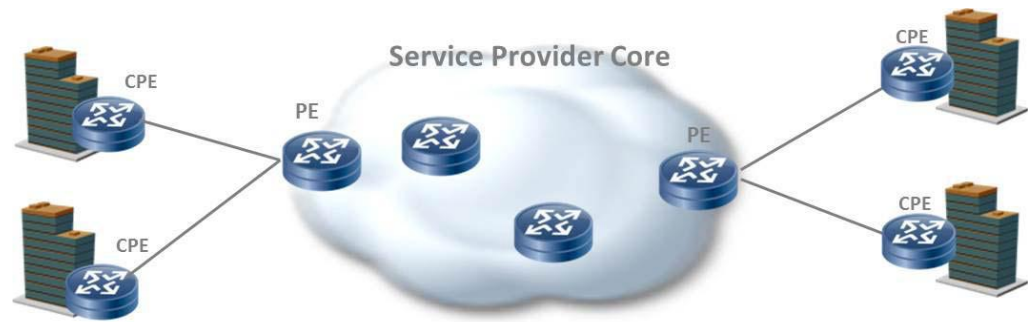


Рис. 1.24 Постачальник послуг без віртуалізації інфраструктури підприємства

Віртуалізація інфраструктури підприємства може включати:

- Віртуалізацію функцій обладнання користувача CPE (vE-CPE) у хмарі постачальника послуг.
- Віртуалізацію функцій порогового маршрутизатора оператора PE (vPE), де функції віртуальних сервісів мережі та функцій ядра PE можуть виконуватись в хмарі постачальника послуг.

Ці два етапи є незалежними і можуть розгортатись окремо. PE маршрутизатори, як правило, надаються великій кількості клієнтів, тоді як маршрутизатор CPE використовується виключно одним клієнтом. Таким чином, економія масштабу, яку можна отримати від віртуалізації CPE є значно більшою, ніж при віртуалізації PE. Отже, віртуалізація CPE є корисною як для користувачів підприємства, так і для постачальників послуг. Віртуалізація PE може відбутись на більш пізньому етапі, щоб завершити перехід до повністю віртуалізованого рішення NFV.

У деяких архітектурах vE-CPE та vPE можуть контролюватися централізованим контролером, який виконує принципи та стандарти SDN (наприклад, OpenFlow).

Постачальник послуг несе відповідальність за розгортання, налаштування, оновлення та керування роботою екземпляру VNF для забезпечення очікуваного рівня обслуговування (SLA) для абонентів VNFaaS.

1.5. Архітектура мережі LTE з віртуалізацією мережевих функцій у гетерогенній хмарній інфраструктурі

Мережеві архітектури, які розробляються на сьогоднішній день, складаються з великої кількості виділених мережевих вузлів, розміщених в різних частинах мережі. Кожен вузол має певний набір функцій і сервісів. Недоліком такої архітектури є те, що вона не гнучка з точки зору додавання нових сервісів. Крім того, вона не забезпечує оптимальну маршрутизацію та ефективне використання мережевих ресурсів. Враховуючи прогнози щодо значного збільшення кількості трафіку у найближчому майбутньому [1], основними проблемами такої архітектури стають масштабованість та економічність.

Одним зі способів підвищення якості надання послуг мобільними операторами, які досліджуються на сьогоднішній день, є використання обчислювальних методів, які дозволяють оптимізувати використання ресурсів мобільної мережі, відповідно до її завантаженості, створювати прогнози щодо кількості трафіку, який передається мережею, а також покращувати показники QoS за рахунок балансування навантаження мережі. Використання обчислювальних методів для оптимізації роботи мобільної мережі стає можливим за рахунок впровадження мережевих функцій у програмне забезпечення, яке може функціонувати на стандартній апаратній платформі[2].

Поширення потужних мобільних пристроїв та збільшення кількості мультимедійних прикладних програмних компонентів, якими користуються абоненти, привело до зростання вимог до пропускної здатності мережі. Згідно з прогнозами компанії Cisco [3], у 2019 році кількість мобільного трафіку, який передається мережею, досягне 24.3 екзабайт на місяць. Це майже у чотири рази більше, у порівнянні з 2016 роком. При цьому очікується, що більшу частину трафіку буде становити передача відео з високою якістю

розподільної здатності - HD-відео, а також 3D-відео. Прогноз компанії Cisco щодо темпів росту мобільного трафіку зображений на Рис. 1.25.



Рис. 1.25 Прогноз Cisco щодо темпів росту мобільного трафіку у екзбайтах на місяць

Послуги, що надаються користувачам мобільного зв'язку є різномірними, та потребують постійного контролю показників якості обслуговування, оскільки різні типи послуг по різному чутливі до затримок та втратив процесі передачі інформаційних потоків. Саме тому для організації гнучкої телекомунікаційної системи оператора необхідно впроваджувати рішення з використанням програмного забезпечення на різних ділянках системи обслуговування. Для забезпечення контролю якості на кожному етапі необхідно контролювати та керувати не тільки процесом обслуговування, але й організації та реорганізації когнітивного ресурсу мережі доступу, що включає в себе радіо ресурс, фізичні/віртуальні обчислювальні ресурси, а також когнітивного ресурсу ядра мережі, який включає в себе фізичні та віртуальні ресурси дата центрів та телекомунікаційних мереж, що використовують для організації роботи розподілених дата центрів.

Отже, необхідно впроваджувати та вдосконалювати системи розподілу ресурсів обслуговування на всіх ділянках мережі мобільного оператора зв'язку.

За базову технологію мобільної системи зв'язку в роботі обрано систему LTE, яка успішно працює вже сьогодні у провідних країнах світу, та поступово до неї наближаються українські компанії. Загальна модель мережі стандарту LTE та її основні інтерфейси показані на Рис. 1.26.

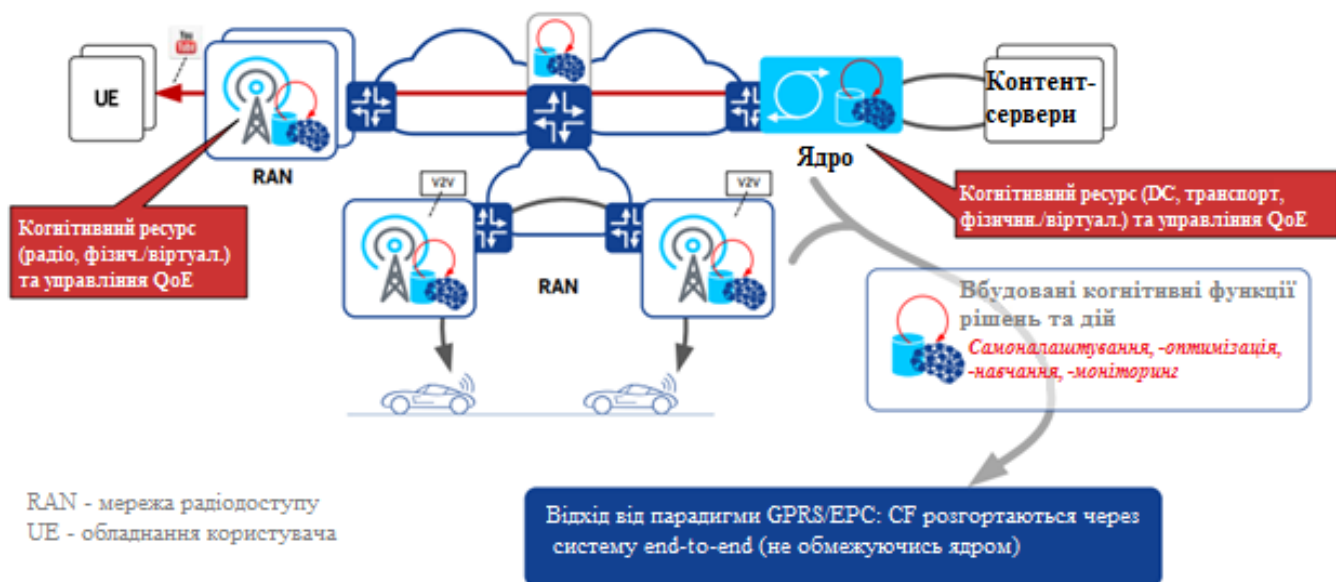


Рис. 1.26 Мережа мобільного оператора зв'язку

Розвиток інформаційних технологій, а саме концепції хмарних обчислень, гібридні хмари, розподілені обчислення, тощо, дало поштовх до розвитку концепцій віртуалізації мережевих функцій, та програмно-керованих мереж. Мобільні мережі п'ятого покоління, які наразі розробляються науковою спільнотою усього світу, передбачають трансформацію системи зв'язку, коли функції підсистем мобільної мережі будуть частково виконуватися як прикладні програмні компоненти.

Враховуючи всі особливості функціонування мереж мобільного зв'язку а також гнучкі можливості які відкриває віртуалізація мережевих функцій та рівень розвитку технології хмарних обчислень, я пропоную архітектуру мережі 5G, яка узагальнює основні тенденції положення та розробки. Архітектура мережі показана на Рис. 1.27.

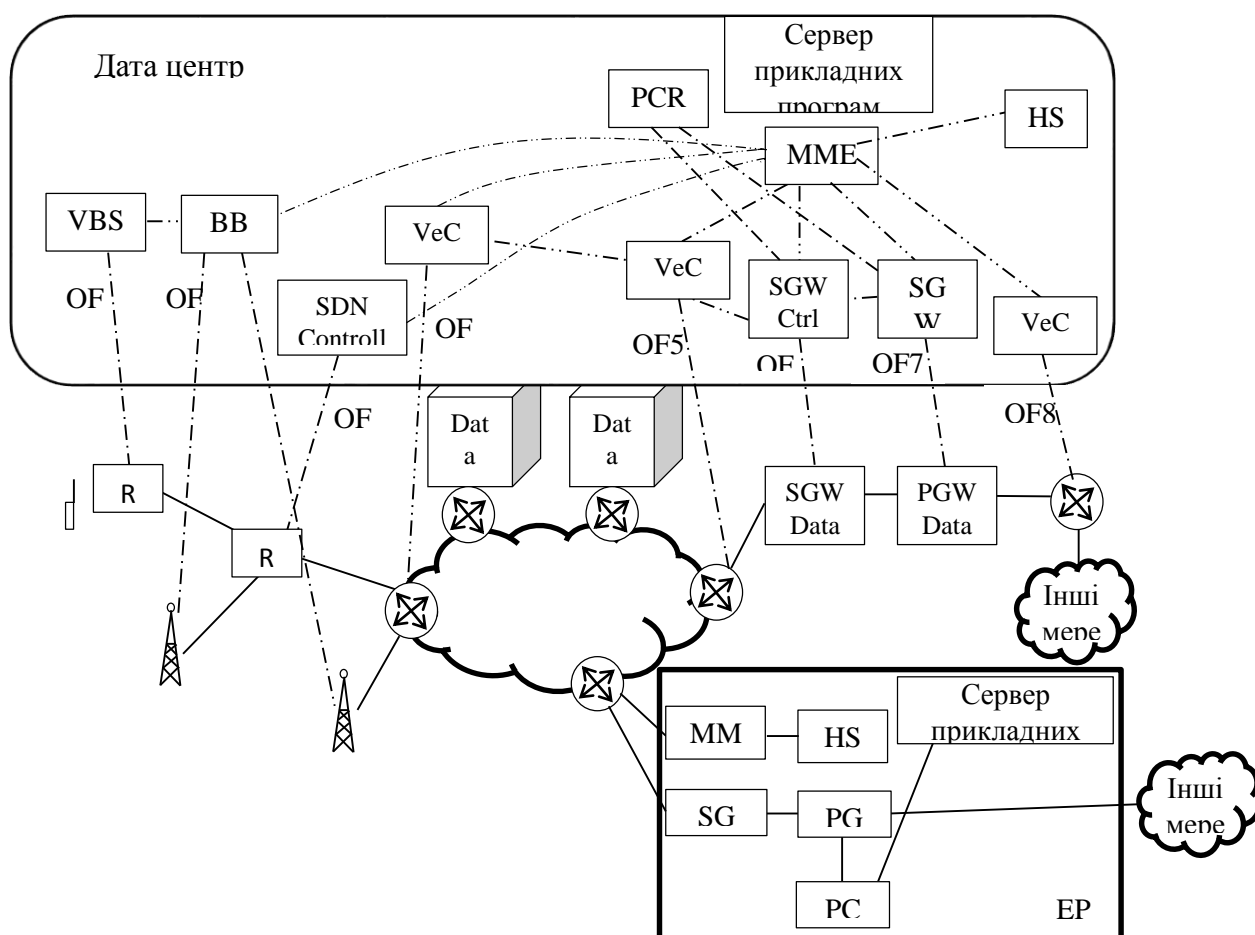


Рис. 1.27 Архітектура мережі оператора зв'язку з віртуалізацією мережевих функцій

Відповідно до специфікації ETSI GS NFV 001 v.1.1.1 (10/2013) всі обчислювальні функції, які супроводжують процес передачі будуть виконуватися у дата центрах із хмарною інфраструктурою, передбачається віртуалізація в тому числі базових станцій, що забезпечить зменшення використаної енергії, пов'язаної з динамічним розподілом ресурсів та балансуванням навантаження. Крім базових станцій для мереж доступу з хмарною організацією ресурсів (Cloud-RAN) передбачається створення ресурсу обробки базових частот (BBU) для об'єднання різноманітних базових станцій як віртуалізованих так і невіртуалізованих у єдине віртуалізоване середовище. Далі віртуалізація мережевих функцій відповідно до специфікації пропонується для маршрутизатора розташованого на межі локальної мережі оператора зв'язку. Маршрутизатор, що виконує

функції класифікації потоків, керування маршрутизацією і забезпечення бар'єрного захисту мережі (Firewall).

Організація віртуальних базових станцій і VeCPE передбачає наявність дата центру поблизу базових станцій і кожного виходу з локальної мережі.

Таким чином, функціонування мережі оператора зв'язку являє собою розподілену територіально мережу дата центрів до кожного з яких підключені канали зв'язку, які доставляють первинну інформацію мобільних абонентів, що вимагає перетворення на найнижчому рівні (тобто сигнал, який потрібно розпізнати і розкодувати на більш високих рівнях MAC, RLC, RRC і PDCP). Таким чином можна бачити, що більшість процесів мережі оператора зв'язку відбувається в дата центрах. Мережі зв'язку стають лише засобом доставки інформаційних повідомлень. В умовах поширення програмно-керованих маршрутизаторів маємо структуру мережі показану на Рис. 1.27.

Тут можна бачити, як мобільний абонент зв'язується з ретранслятором R1, який перетворює радіосигнал в оптичний, потім сигнал доходить до ретранслятора R2, керованого SDN контролером, який так само знаходиться в дата центрі. Потрапляючи в дата центр, сигнал обробляється віртуальною базовою станцією. Далі згідно з технологією LTE потік прямує в ядро оператора для подальшої обробки.

Підсистема BBU (блок формування модулюючих сигналів) заснована на технології програмно-конфігурованих мереж/віртуалізованих функцій мережі, це система, яка підтримує роботу не лише роботу віртуальних базових станцій, але працює для гібридних рішень 2G/3G/4G/Pre5G.

Слід зазначити, що обслуговування в ядрі визначає подальший напрямок потоку даних. Якщо потік спрямований у внутрішню мережу оператора, тоді тут же в дата центрі він направляється на обслуговування у відповідну віртуальну базову станцію, і далі направляється до абонента через ретранслятори R2 і R1. Якщо його призначення лежить за межами локальної мережі оператора зв'язку, тоді потік прямує на віртуальний маршрутизатор

кордону локальної мережі, який знаходиться тут же в дата центрі. А після обслуговування потік прямує в зовнішні мережі.

Таким чином, будуть виглядати мережі наступних поколінь. Однак якщо подивитися на елемент дата центр, то стає зрозуміло, що він об'єднує групу дата центрів з'єднану через захищену мережу в єдиний логічний простір обслуговування. Забезпечення якісного обслуговування кінцевих абонентів істотно залежить від організації процесів в такому гетерогенном дата центрі, побудованому виходячи з концепції хмарних обчислень.

Згідно з рекомендацією ITU-T Y.3500 хмарні обчислення - це парадигма забезпечення мережевого доступу до масштабованого і гнучкого набору спільно використовуваних фізичних і віртуальних ресурсів з наданням та адмініструванням ресурсів на підставі самообслуговування за вимогою.

Телекомунікаційна структура описаного дата центру в якому виконується обслуговування групи функціональних блоків показаних на Рис. 1.27, представлена на Рис. 1.28, де можна бачити транспортну мережу і підключені до неї дата центри, що утворюють єдиний віртуалізований простір.

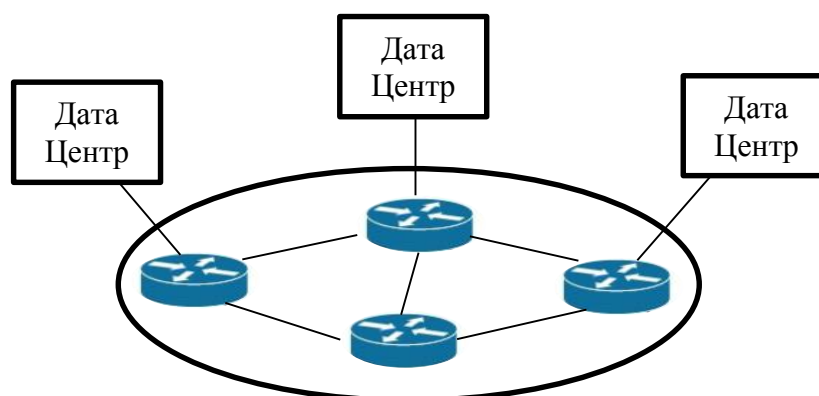


Рис. 1.28 Телекомунікаційна схема розміщення дата центрів на транспортній мережі.

У мережах наведених на Рис. 1.27 ефективність обслуговування абонентів залежить від ефективності організації роботи дата центра. У даному архітектурному рішенні під дата центром розуміється складана організаційно-технічна сукупність обчислювальних та телекомунікаційних

системи які забезпечують безперебійну роботу інфраструктури NFV. Ефективність роботи дата центру залежить:

- від вибору фізичних дата центрів, які увійдуть до складу розподіленої структури дата центрів;
- від того, як буде вирішено питання з розміщенням мережевих функцій у інфраструктурі розподілених дата центрів;
- від організації потоків між віртуалізованими сутностями;
- від виділення ресурсів для обслуговування віртуалізованих сутностей.

В роботі будуть досліджуватися особливості ефективної роботи мобільних мереж 5-го покоління, та способи забезпечення показників якості обслуговування кінцевих користувачів.

Висновки

Функціонування сучасного інформаційно-телекомунікаційного середовища можна охарактеризувати як роботу складної системи, що обслуговує різні типи сервісів, які надаються не лише кінцевим абонентам, але й програмно-керованим автоматизованим системам, забезпечується різноманітним обладнанням, робота якого суттєво залежить від програмного забезпечення систем керування процесом обслуговування. Системи керування процесом обслуговування сервісів набувають все більшого розвитку. Наразі для забезпечення процесу керування розробляються стандарти, моделі та підходи із застосуванням технології хмарних обчислень.

При трансформації телекомунікаційних систем у інформаційно-телекомунікаційні з використанням хмарних сервісів виникає необхідність у методологічному базисі організації ефективної роботи інформаційно-телекомунікаційного середовища до вимог та особливостей процесу обслуговування ТС-гібридних сервісів.

В розділі визначено та систематизовано підходи щодо організації процесу обслуговування у інформаційно-телекомунікаційних мережах, з

частковою віртуалізацією мережевих функцій. Окреслено наступні невирішені проблеми:

- потрібно контролювати затримки у віртуалізованих мережевих вузлах, де інтенсивність обслуговування залежить від обчислювальних ресурсів – потребує модифікації PCRF;

- потрібно мати можливість реконфігурації ресурсів обслуговування – потреба у розробці нової системи керування ресурсами обслуговування гібридних телекомунікаційних сервісів у гетерогенній системі даних центрів

На основі проведених досліджень сформульована актуальна наукова проблема розробки методологічного базису для керування процесом обслуговування гібридних телекомунікаційних сервісів у гетерогенному телекомунікаційному середовищі, що забезпечить обслуговування кінцевих користувачів на заданому рівні якості та дозволить керувати телекомунікаційними та обчислювальними ресурсами, які забезпечують їх обслуговування.

Забезпечення якості послуг для кінцевих користувачів мереж зв'язку залежить від контролю якості на всіх етапах надання кінцевих послуг. Сьогодні, завдяки динамічно змінній структурі послуг, що надаються кінцевим споживачам, постійно змінюються вимоги до якості показників послуг та зростання обсягів трафіку, виникає потреба у високорозвинених системах зв'язку, які б задовольняли потреби кінцевих користувачів. Цей документ описує методи підвищення якості обслуговування гібридних телекомунікаційних мереж, які передбачають систематичне застосування методів моніторингу та керування процесом формування потоків для обслуговування та розподілу сервісних ресурсів

РОЗДІЛ 2

МОДЕЛІ КОНТРОЛЮ ЯКОСТІ ОБСЛУГОВУВАННЯ КІНЦЕВИХ КОРИСТУВАЧІВ

2.1. Моделі контролю якості обслуговування

Забезпечення якості обслуговування для телекомунікаційних систем включає в себе методи контролю та керування роботою телекомунікаційної системи. На сьогоднішній день розроблено безліч механізмів адаптивного реагування на зниження якості обслуговування. Можна виділити наступні групи механізмів:

1. Контроль показників обслуговування черг у вузлах зв'язку.
2. Контроль завантаженості мережі.
3. Управління організацією подій для диференційованого обслуговування мультисервісних потоків
4. Контроль параметрів потоків абонентських даних
5. Методи зворотного зв'язку - механізми попередження джерел даних про можливі перевантаження вибраного напрямку зв'язку.
6. Методи інжинірингу трафіку для планування рівномірного завантаження ресурсів мережі.

Робота системи контролю якості обслуговування для телекомунікаційної системи з віртуалізацією мережевих функцій пов'язана з контролем та керуванням телекомунікаційної системи, а також з контролем роботи інформаційно-обчислювальної системи. Інформаційно-обчислювальна система забезпечення мобільного зв'язку, розгортається на базі групи дата центрів територіально розподілених та об'єднаних у єдиний обчислювальний простір за технологією гібридна хмарна система.

Наразі досліджуються два основні підходи до віртуалізації:

- Віртуалізація функцій керування мережею, що передбачає виконання завдань пов'язаних з обслуговуванням службових потоків у відокремлених обчислювальних системах розміщених у дата центрах із гетерогенною

хмарною інфраструктурою. Всі потоки даних проходять через телекомунікаційні системи, їх не перенаправляють до хмар, відповідно не створюють надлишкове навантаження на телекомунікаційну систему.

- Віртуалізація всіх функцій телекомунікаційної системи, коли для обслуговування телекомунікаційного сервісу його направляють до дата центру, приклад віртуальна базова станція з віртуалізацією функції кодування/декодування, яка передбачає виконання операцій не лише з потоками керування, але й з потоками даних.

Контроль якості обслуговування передбачає перевірку відповідності параметрів обслуговування заданим пороговим значенням показників якості обслуговування на кожному етапі обслуговування сервісу. Порогові значення параметрів якості обслуговування користувачів зберігаються у базах підсистеми PCC (Policy and Charging Control), контроль параметрів обслуговування здійснюється для кожного користувача індивідуально відповідно до договору на обслуговування. Основними класичними, в термінах теорії масового обслуговування, параметрами QoS (Quality of service) є затримка у вузлі обслуговування, втрати заявок через перевищення часу очікування на обслуговування, втрати заявок через перевантаження черг на обслуговуючих пристроях. Саме ці параметри досліджують.

На сьогоднішній день розроблено групу стандартів та специфікацій [218, 219], в яких описано підходи до контролю якості обслуговування у системах мобільного зв'язку 3G, 4G. Основною підсистемою яка перевіряє відповідність поточних показників якості заявленим у PCC показникам є PCRF (Policy and Charging Rules Function). В стандартах описані основні функції та процедури забезпечення контролю та керування якістю, а також взаємодія підсистем мобільного зв'язку.

Впровадження підходів віртуалізації змінить функціонування системи забезпечення якості обслуговування.

При реалізації підходу з віртуалізацією системи керування (Рис. 2.1) основною задачею є забезпечення вимог до якості функціонування системи

керування, а саме своєчасність прийняття керівних рішень. Моніторинг роботи системи мобільного зв'язку здійснює підсистема SGSN. Основні параметри, які контролюються описано у п. 2.2.3. При виявленні невідповідності показників доступності послуг пороговим значенням, виникає потреба у оптимізації роботи системи керування. Ефективність функціонування віртуалізованої системи керування залежить від ефективності організації обчислювального процесу.

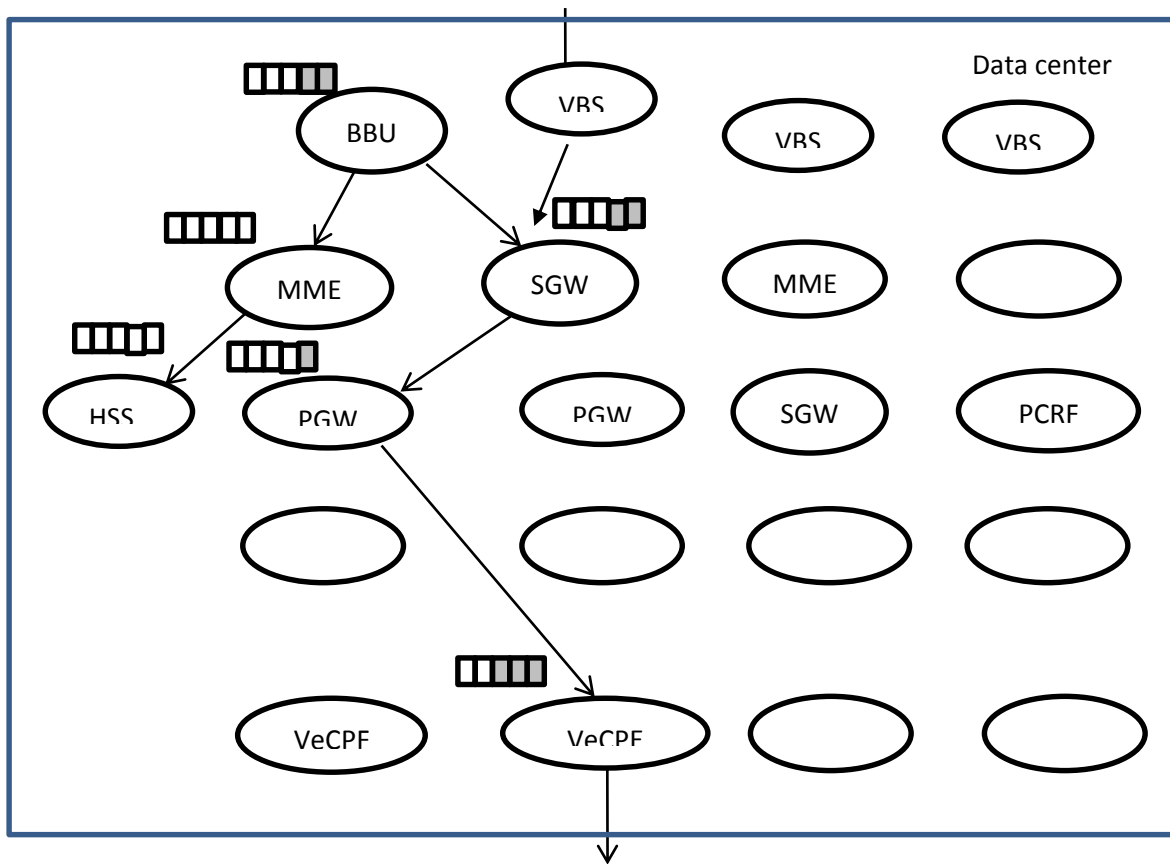


Рис. 2.1 Обслуговування потоків в Інфраструктурі NFV

Розроблено також архітектурну модель логічної структури віртуальної мережі, на базі якої розгортається інфраструктура NFV. Кількість вузлів у віртуальній мережі а також кількість віртуалізованих сутностей залежить від структури вхідних потоків від абонентів та розташування точок їх входу (надходження) до системи обслуговування. *Особливістю віртуалізованого середовища* є гнучкість та можливість динамічної зміни структури мережі, місце обслуговування потоків, та кількість NFV-ресурсів, залучених до

процесу обслуговування. Основним засобом контролю якості обслуговування гібридних сервісів є контроль показників обслуговування

2.1.1. Вимоги до якості в 5G мережі

В процесі еволюції механізму управління QoS в GSM/UMTS/LTE мережах сталася міграція управління QoS з рівня користувацького обладнання до керування на рівні мережі. Цей підхід також збережеться в мережах 5G.

Механізми керування QoS в мережах 5G повинні забезпечити пріоритет відео і VoIP трафіку над іншими сервісами. Сервіс потокового відео без буферизації дуже чутливий до затримок в мережі, тому одним з найбільш важливих параметрів, який визначає вимоги до QoS є загальний час затримки пакетів (packet delay budget - PDB). У Таблиця 2.1 наведені вимоги до затримки в 3G/4G/5G мережах.

Таблиця 2.1

Вимоги до затримки в 3G/4G/5G мережах

Терміни QoS	Запланована затримка пакетів (мс)		
	3G	4G	5G
Без гарантування якості	Не визначено	100-300	Не визначено
З гарантованою якістю	100-280	50-300	1

Ці дані показують, що з ростом покоління мобільної мережі, вимоги до нижньої межі затримки даних збільшуються. Також аналіз вимог до загальної мережевої затримки 5G показав, що вона повинна бути менше 1 мс.

На Рис. 2.2 представлено порівняння вимог до затримки на рівні управління та рівні користувача для сигнального та абонентського трафіків відповідно.

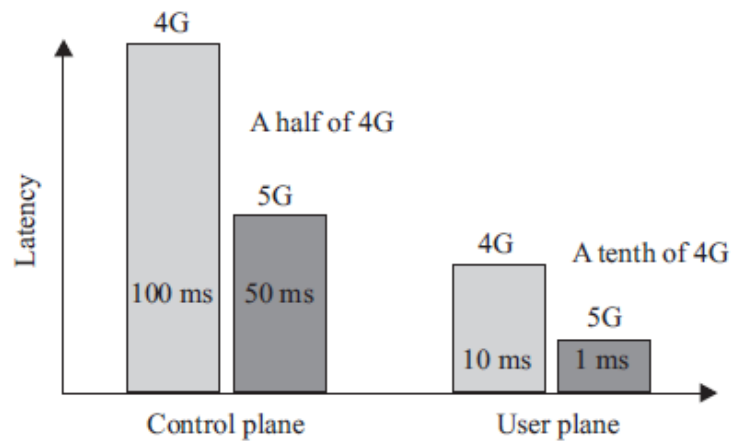


Рис. 2.2 Порівняння вимог до затримки на рівні управління та рівні користувача

На Рис. 2.2 видно, що вимоги до мережі 5G будуть в два рази жорсткішими для сигнального трафіку і в 10 разів жорсткішими для абонентського трафіку.

Іншим параметром є частка втрачених пакетів через помилки при отриманні пакетів даних (IP-packet Error Rate).

Значення цього параметра, що визначає вимоги максимального числа втрат IP пакетів для широкомовного відео за допомогою мобільних мереж 3G/4G/5G, наведені в Таблиця 2.2

Таблиця 2.2

Вимоги до кількості втрачених пакетів для широкомовного відео в 3G/4G/5G мережах

Терміни QoS	Запланована затримка пакетів (мс)		
	3G	4G	5G
Без гарантування якості	10^{-2}	10^{-3}	10^{-4}
З гарантованою якістю	10^{-2}	10^{-6}	10^{-7}

Для інших сервісів якість також буде визначатися часткою втрачених пакетів в мережах 3G/4G/5G. Умови обслуговування абонентських пристроїв будуть визначатися в обох випадках: з гарантованою якістю обслуговування і

без гарантованої якості. Вимоги до коефіцієнта втрат пакетів для інших послуг наведені в Таблиця 2.3.

Таблиця 2.3

Вимоги до частки втрачених пакетів для інших сервісів в 3G/4G/5G мережах

Терміни QoS	Коефіцієнт втрат пакетів (Packet Error Loss Rate)			
	SDTV	HDTV	4k UHD	8k UHD
Покоління мобільних мереж	3G/4G	4G	4G	5G
Широкомовне відео з гарантованою якістю	10^{-6}	10^{-7}	10^{-8}	10^{-9}

Розвиток концепції NFV призведе до віртуалізації функцій управління якістю, які можуть бути розділені на 2 складові: управління (Cloud QoS management function - CQMF) і контролю (Cloud QoS control function - CQCF), які показані на Рис. 2.3.

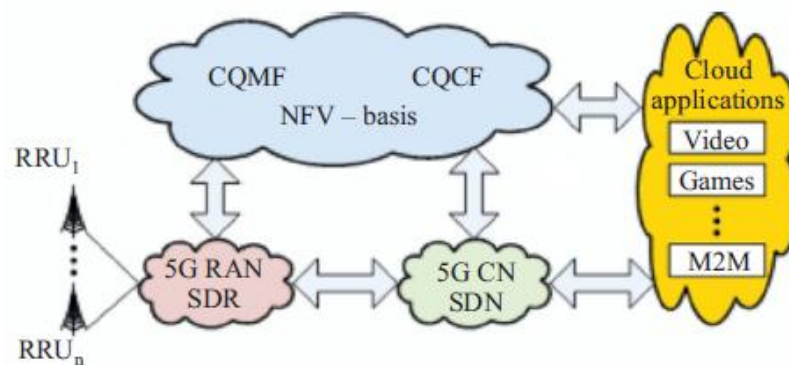


Рис. 2.3 Основні складові функцій управління якістю

Функція CQCF забезпечує контроль потоків трафіку в режимі реального часу на основі QoS, встановленого під час з'єднання. Основні механізми управління QoS включають управління профілями трафіку, планування і управління потоками даних.

Функція CQMF забезпечує підтримку QoS в мережі 5G відповідно до SLA, а також забезпечує моніторинг, технічне обслуговування, аналіз і масштабування QoS.

Реалізація алгоритмів пріоритетності трафіку в мережах 5G буде ґрунтуватися на процедурах класифікації трафіку з акцентом на пріоритети відео-трафіку і M2M трафіку. Класифікація процедур трафіку повинна бути проведена з урахуванням можливостей адаптації: характеристики трафіку будуть динамічно змінюватися з появою нових додатків, як в M2M області, так і в області відео.

2.2. Контроль якості у системі з віртуалізованою системою керування

2.2.1. Забезпечення якості в мережах LTE без віртуалізації

Оркестрація та менеджмент (O&M) базової станції UMTS (Node B) Node B розділена на дві частини: O&M, пов'язана з фактичною реалізацією базової станції, позначена як конкретна реалізація (Implementation Specific) O&M, та O&M, що впливає на ресурси передачі трафіку в базовій станції, яка управляється контролером радіомережі, позначена як логічна O&M. Архітектура контролера мережі (RNC) з інтерфейсами O&M показана на Рис. 2.4.

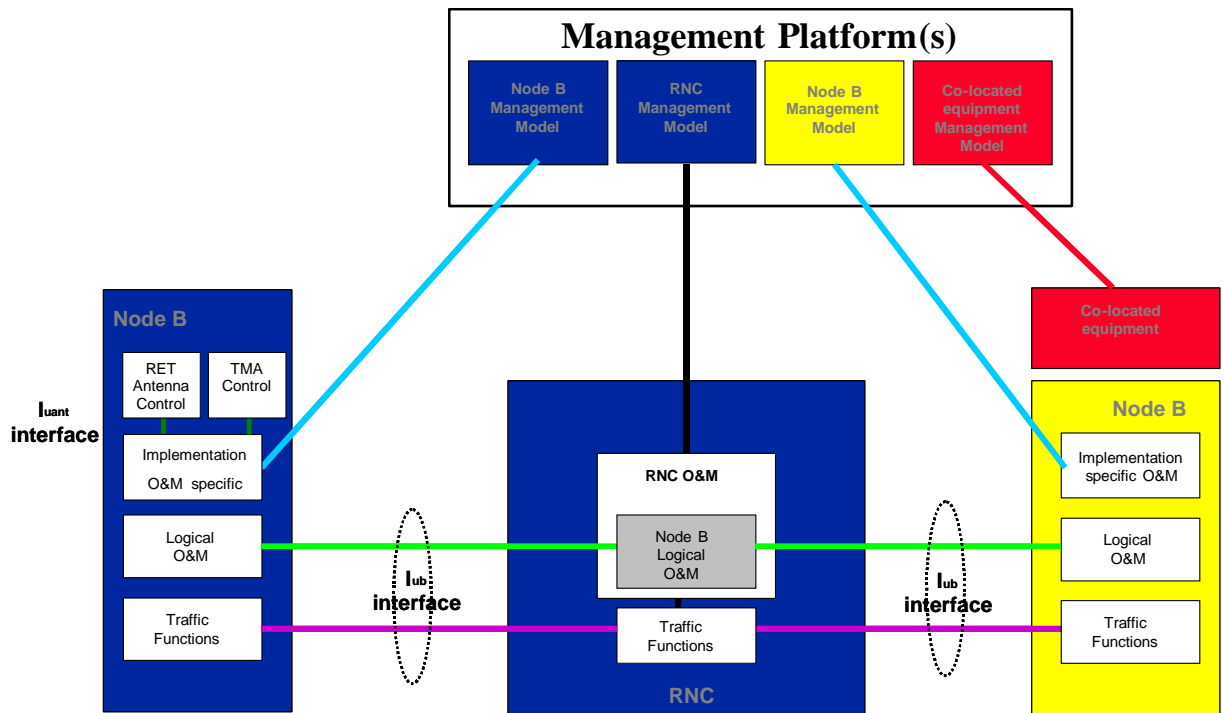


Рис. 2.4 Архітектура контролера мережі з інтерфейсами O&M

На функції конкретної реалізації O&M значною мірою впливає реалізація базової станції (NodeB), як з точки зору її апаратних компонентів, так і з точки зору управління компонентами програмного забезпечення. Тому функції є залежними від реалізації та виконуватися між базовою станцією та системою управління.

Як варіант транспортного рішення для конкретної реалізації O&M - це маршрут від базової станції до системи управління через контролер мережі RNC. У цьому випадку інтерфейс конкретної реалізації O&M для та інтерфейс Iub мають одного і того ж фізичного каналу, а TS 25.442 [4] визначає функцію маршрутизації та транспортного каналу для цього сценарію. Розгортання маршрутизації через RNC в UTRAN є обов'язковим. Там, де потрібна сигналізація між спільно розміщеним обладнанням та його системою управління, вона може бути перенесена на той самий канал, що і конкретна реалізація O&M.

Логічна O&M - сигналізація, пов'язана з контролем логічних ресурсів (каналів, сот, ...), що належать контролеру мережі RNC, але фізично реалізовані у базовій станції. Контролер радіомережі контролює ці логічні ресурси. Ряд процедур O&M, фізично реалізованих у Node B, впливають на логічні ресурси, і тому вимагають обміну інформацією між RNC та Node B. Усі повідомлення, необхідні для підтримки цього інформаційного обміну, класифікуються як логічні O&M, які є невід'ємною частиною сигнального протоколу NBAP (Node B Application Part).

2.2.2. Вимоги до NFV і SDN інтеграції в мобільних мережах

Основні вимоги до гетерогенних мереж поряд зі скороченням витрат операторів полягають в задоволенні необхідного значення показника доступності, який в свою чергу включає показники:

- відмовостійкості;
- підвищеної продуктивності;
- низькою затримки.

Крім цього потрібна наявність вбудованого моніторингу, забезпечення чіткого шляху міграції, сумісності з успадкованими системами, забезпечення актуального переліку послуг, що надаються.

Вимоги до рівня доступності сучасних мобільних мереж розглянуті в роботі [3] і автори відзначають, що значення рівня доступності має бути близько 99,999%, цей показник повинен бути збережений або покращений.

Інтеграція SDN та віртуалізація мережевих функцій зводить до мінімуму зміни в мережевих елементах, забезпечуючи тим самим плавний перехід на основі потреб оператора. Безпека має першочергове значення для мобільних мереж і її слід розглядати для всіх рівнів, мережевих функцій, а також фізичних і віртуальних елементів. Починаючи від контролера SDN, який має доступ до всієї мережевої архітектури, закінчуючи вузлом, який виконує

функції мережі, система повинна гарантувати критичний рівень безпеки і високої доступності.

Ще однією важливою вимогою є моніторинг мережі. Моніторинг мережі полегшує перевірку і підтвердження Service Level Agreements (SLA), продуктивності (якість обслуговування - QoS), пошуку та усунення несправностей, а також оцінку та оптимізацію використання ресурсів. З одного боку, віртуалізація мережевих функцій встановлює нові вимоги до моніторингу мобільного мережі, але з іншого боку також надає засоби для реалізації передових рішень моніторингу мережі. NFV/SDN дозволяє інтегрувати хмарні інфраструктури, що забезпечують більш високу ступінь свободи щодо розміщення точок вимірювання та гнучкого управління потоками трафіку. Передове і ефективне рішення моніторингу QoS має включати як розподілену (на основі SDN/NFV) систему вимірювання QoS, так і централізовану систему оцінки.

Налаштування сервісів і оптимізація – ще одні вимоги для забезпечення доступності ресурсів. Це може бути зроблено за допомогою оркестратора. Ця вимога може бути розгорнута в SDN мережах, використовуючи додатки управління, які мають повне уявлення про конфігурацію мережі. Разом з інформацією про статус від систем моніторингу мережі та збору даних, це дозволяє оркестратору мобільної мережі оптимізувати обслуговування (наприклад, затримку) і/або використовувати менше ресурсів, ніж традиційні мережі. Оркестратор може контролювати кілька мережевих елементів через додатки управління. Це дозволяє впроваджувати нові послуги вносячи зміни в оркестратор, тоді як в традиційних мережах для підтримки нових сервісів необхідно оновлювати все обладнання. Мереж на базі SDN повинні співіснувати з наявною архітектурою мережі. Для того, щоб задіяти потенціал SDN, потрібна взаємодія цих двох рішень, наприклад, шляхом введення рівня абстракції і автоматизації в застарілу мережеву частину.

2.2.3. Показники якості послуг передачі даних

В системах мобільного зв'язку третього покоління функції контролю параметрів доступності сервісів виконувала підсистема SGSN, в ядрі EPC ці функції виконує підсистема PCRF.

На Рис. 2.5 зображена архітектура мережі передачі даних 2G/3G. Основним мережевим вузлом в цій архітектурі є вузол підтримки послуг GPRS (Serving GPRS Support Node, SGSN), який відстежує розташування абонентських терміналів передачі пакетних даних, забезпечує захисні функції і контроль доступу. SGSN з'єднується з контролером базових станцій BSC, використовуючи Frame Relay - мережу комутації фреймів (пакетів другого, каналного рівня). Вузол шлюзовий підтримки GPRS (Gateway GPRS Support Node, GGSN) взаємодіє з зовнішніми мережами пакетної передачі даних (Packet Data Networks, PDNs), забезпечуючи передачу даних до/від мобільних терміналів. GGSN пов'язаний із зовнішніми мережами з пакетним перемиканням і з'єднується з вузлами SGSN через мережі, які використовують протокол IP. Керуючий блок пакетної комутації (Packet Controller Unit, PCU) забезпечує послуги пакетної радіопередачі на область охоплення BSC. Для з'єднання вузлів SGSN і GGSN з іншими елементами мережі, що забезпечують глобальну пакетну передачу даних, може додаватися кілька нових інтерфейсів (з позначенням G *, де * - символ, що визначає конкретний інтерфейс).

Розглянемо основні причини збою в роботі системи передачі даних. Однією з найбільш поширених причин є помилка, яка відправляється з боку GGSN на Gn інтерфейс. Ця помилка широко поширена в GTP повідомленнях інформаційного елемента.

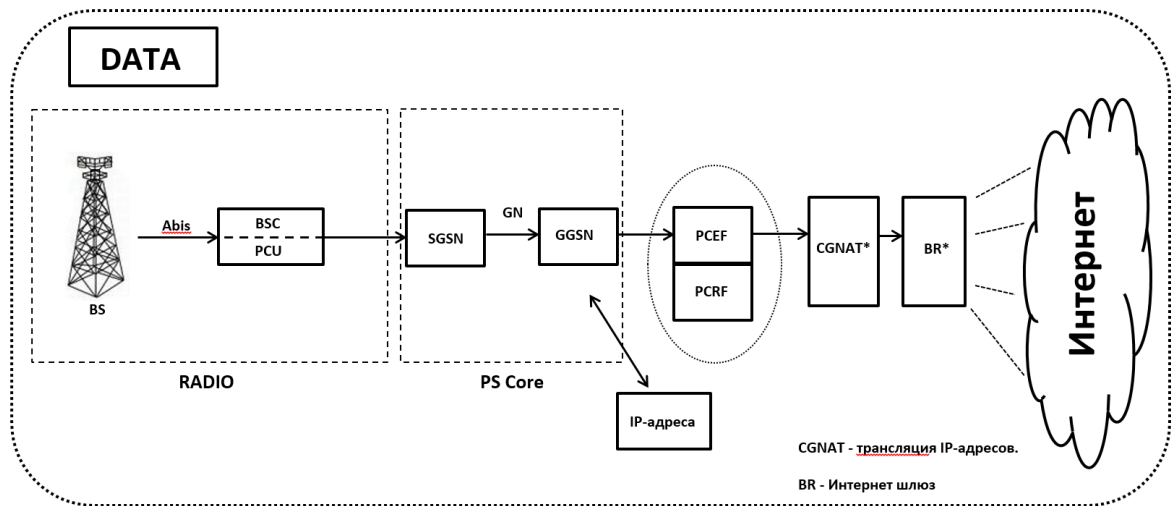


Рис. 2.5 Архітектура мережі передачі даних 2G/3G

PDP Context - набір даних про абонента, що здійснив процедуру GPRS Attach (процедуру аутентифікації і надання доступу в інтернет), який зберігається як на стороні SGSN, так і на стороні терміналу абонента, зокрема в цей набір даних входить профіль, що забезпечує певний рівень якості обслуговування абонента - QoS, призначена абоненту адреса в мережі, деякі дані про тарифікацію абонента.

Розглянемо основні показники якості сервісу передачі даних

1. Доступність

- Відсоток успішно виконаних процедур активації PDP context, ініційованих MS 2G/3G (PDP Context Activation Success Rate)

$$PDPctx_act_SR = \frac{PDPctx_act_acc}{PDPctx_act_req} * 100, \text{ де}$$

PDPctx_act_acc - число успішно виконаних процедур активації PDP-контексту, ініційованих MS (таймаут = 150с, після чого спроба вважається неуспішною).

PDPctx_act_req - число запитів на виконання процедури активації PDP-контексту, ініційованих MS.

- Затримка часу встановлення TCP-сесії для сервісу Web Browsing

$$WebBrowsin gLatency =$$

$$\frac{1}{N} \sum (WBST - WBRT)_i, \text{ де}$$

WBST(WebBrowsingSuccessTime) - час встановлення TCP-сесії для сервісу Web Browsing;

WBRT(WebBrowsingRequestTime) - час запиту встановлення TCP-сесії для сервісу Web Browsing і -ий запит встановлення TCP-сесії

N - сумарна кількість запитів за інтервал спостереження.

Відсоток блокувань в ЧНН в режимі передачі даних через перевантаження на 2G/3G (Відношення між кількістю відмов у виділенні ресурсів для передачі даних через перевантаження і загальною кількістю запитів на виділення ресурсів в ЧНН 2G/3G (Connection Block Rate).

$$\text{ConnectionBlockRate} = (PS_BlockRate2G * \text{Traffic2G} + PS_BlockRate3G * \text{Traffic3G}) / (\text{Traffic2G} + \text{Traffic3G})$$

PS_BlockRate2G - відсоток блокувань в ЧНН в режимі передачі даних через перевантаження для мережі 2G

PS_BlockRate3G - відсоток блокувань в ЧНН в режимі передачі даних через перевантаження для мережі 3G

Traffic2G - пакетний трафік в мережі 2G в ЧНН, MB

Traffic3G - пакетний трафік в мережі 3G в ЧНН, MB

- Відсоток успішно виконаних спроб реєстрації мобільної станції в мережі пакетної передачі даних 2G/3G (PS Attach SR).

$$PS_Attach_SR2G/3G = (PS_Attach_SR2G * \text{Traffic2G} + PS_Attach_SR3G * \text{Traffic3G}) / (\text{Traffic2G} + \text{Traffic3G})$$

PS_Attach_SR2G - відсоток успішно виконаних спроб реєстрації мобільної станції в мережі пакетної передачі даних 2G;

PS_Attach_SR3G - відсоток успішно виконаних спроб реєстрації мобільної станції в мережі пакетної передачі даних 3G;

Traffic2G - середньодобовий пакетний трафік в мережі 2G, MB;

Traffic3G - середньодобовий пакетний трафік в мережі 3G, MB;

- Відсоток успішно виконаних процедур активації PDP context, ініційованих MS 2G/3G (PDP Context Activation Success Rate)

$$\begin{aligned} PDPctx_act_SR2G/3G = \\ (PDPctx_act_SR2G * Traffic2G + \\ PDPctx_act_SR3G * Traffic3G) / \\ (Traffic2G + Traffic3G) \end{aligned}$$

PDPctx_act_SR2G - відсоток успішно виконаних процедур активації PDP context, ініційованих MS в 2G

PDPctx_act_SR3G - відсоток успішно виконаних процедур активації PDP context, ініційованих MS в 3G

Traffic2G - середньодобовий пакетний трафік в мережі 2G, MB

Traffic3G - середньодобовий пакетний трафік в мережі 3G, MB

- Стабільність роботи центральних систем пропуску data-трафіку

$$data_traf_stability = 1 - \frac{t_{traf_unstable}}{t}$$

ttraf_unstable - час нестабільної роботи систем пропуску трафіку вважається період, протягом якого рівень трафіку на PS-Core був нижче більш ніж на 20% рівня відповідного періоду минулого тижня. Показник стабільності роботи центральних систем пропуску трафіку є відношенням часу, протягом якого системи працювали стабільно, до загального часу періоду *t*.

2. Цілісність

- Частка успішних тестів WebBrowsing з часом завантаження сторінки Kepler розміром 800 Кбайт не більше 130 секунд,

$$\begin{aligned} Activities_WBtime < 130sec = \\ \frac{Activities_WBtime < 130sec}{TotalActivities} * 100\% , \text{ де} \end{aligned}$$

Activities_WBtime < 130sec - кількість абонентських сесій з WebBrowsing Time < 130sec, *TotalActivities* - загальна кількість абонентських сесій за період вимірювань.

- Середня швидкість передачі даних на одну абонентську активність для сервісу Web Browsing

$$WebBrowsingAverageSpeed = \frac{1}{N} \sum WebBrowsingSpeed_i, \text{ де}$$

i - i -а абонентська активність WebBrowsing-a

N - сумарна кількість запитів сервісу WebBrowsing.

- Частка абонентських сесій, які отримують дані при використанні VS із середньою швидкістю понад 400 кбіт/с

$$Activities_VS > 400kbps = \frac{Activities_VS > 400kbps}{TotalActivities} * 100\%, \text{ де}$$

$Activities_VS > 400kbps$ - кількість абонентських сесій з VideoStreaming Speed > 400kbps,

$TotalActivities$ - загальна кількість абонентських сесій за період вимірювань.

2.2.4. Аналіз методів забезпечення параметрів QoS в сервісно-орієнтованій архітектурі LTE

Концепція системи QoS для мереж UMTS мобільного зв'язку 3-го покоління визначена в специфікації TS 23.107, і використовується також для мереж LTE 4-го покоління [34, 35, 36].

На Рис. 2.6 розглянуто архітектуру системи якості обслуговування і передавання послуг у рамках такої системи, для випадку, коли зв'язок здійснюється між кінцевим обладнанням (КО), підключеним до призначеного для користувача терміналу мобільної мережі, і термінальним обладнанням, розташованим в зовнішній пакетній мережі.

Аналогічно поняттю наскрізного каналу вводиться поняття *наскрізної послуги* (end — to — end service) як послідовність дій між двома кінцевими користувачами і, відповідно, частин послуг — по їх відношенню до визначених мережевим складовим: в локальному каналі "КО — призначений

для користувача термінал" (Terminal Equipment / Mobile Terminal local Bearer Service), в каналі мережі LTE (LTE Bearer Service), в зовнішньому каналі (External Bearer Service). Таким чином, виникає багаторівнева взаємодія при передачі послуги в різних мережевих вузлах і на різних рівнях.



Рис. 2.6 Архітектура системи якості обслуговування [55]

Передача послуги з мережі LTE розглядається, відповідно до мережевої архітектури, окремо в мережі радіодоступу (Radio Access Bearer Service), де забезпечується конфіденційне передавання призначених для користувача даних або із заздалегідь вибраним або встановленим за замовчуванням рівнем якості обслуговування, і у базовій пакетній мережі (Core Network Bearer Service), що також може підтримувати різну якість обслуговування.

Послугу в мережі радіодоступу реалізують двома частинами: в радіоканалі (Radio Bearer Service) і в механізмі радіодоступу (Access Bearer

Service). Реалізація послуги в радіоканалі містить усі аспекти, що стосуються передавання даних по радіоінтерфейсу, включаючи сегментацію і повторне збирання призначених для користувача пакетів. Крім того, на фізичному рівні (Physical Radio Bearer Service) здійснюється управління підпотокм призначених для користувача даних. Механізм радіодоступу забезпечує на фізичному рівні (Physical Bearer Service) передавання даних між мережею радіодоступу і базовою мережею.

Нарешті, проходження послуги в "магістральному" каналі (Backbone Network Bearer Service) розглядається у функціональній сукупності рівнів 1 і 2 і призначених вимог якості обслуговування.

Основні функції мережі LTE, що належать до керування якістю обслуговування в площині користувача

Функція відображення (MF, Mapping Function) забезпечує маркування кожного призначеного для передавання пакету даних відповідними параметрами QoS.

Функція класифікації (CF, Classification Function) призначена для виставлення пакетам параметрів QoS, призначених для певного АТ, у тому випадку, якщо для цього АТ в мережі встановлено декілька каналів передавання послуг.

Функція управління ресурсами (RMF, Resource Manager Function) розподіляє доступні ресурси між послугами відповідно до параметрів QoS.

Функція узгодження (очищення) трафіку (TCF, Traffic Conditioner Function) забезпечує узгодження між потоком призначених для користувача даних і встановленим рівнем якості обслуговування. Ті пакети даних, які не відповідають виставленим параметрам QoS, будуть відкинуті або помічені як невідповідні для наступного відкидання після накопичення.

На Рис. 2.7 показана взаємодія функцій керування якістю обслуговування в призначеній для користувача площині [25].

Функція класифікації, реалізована в абонентському терміналі АТ і сигнальному шлюзі СШ, призначає пакети даних, отримані із зовнішнього

(чи локального) каналу в послугу мережі LTE з відповідними параметрами QoS. Функція узгодження трафіку, при необхідності, забезпечує узгодження призначеного для користувача потоку у висхідному (у АТ) і низхідному (у СШ) напрямках зі встановленими параметрами QoS. Далі функція відображення забезпечує кожен пакет даних спеціальним QoS-індикатором, відправляючи того в мережу, що вимагає виділення відповідних ресурсів — за це відповідальна функція управління ресурсами, реалізована в кожному мережевому вузлі.

У площині управління зосереджені функції, необхідні для реалізації механізмів управління і контролю.

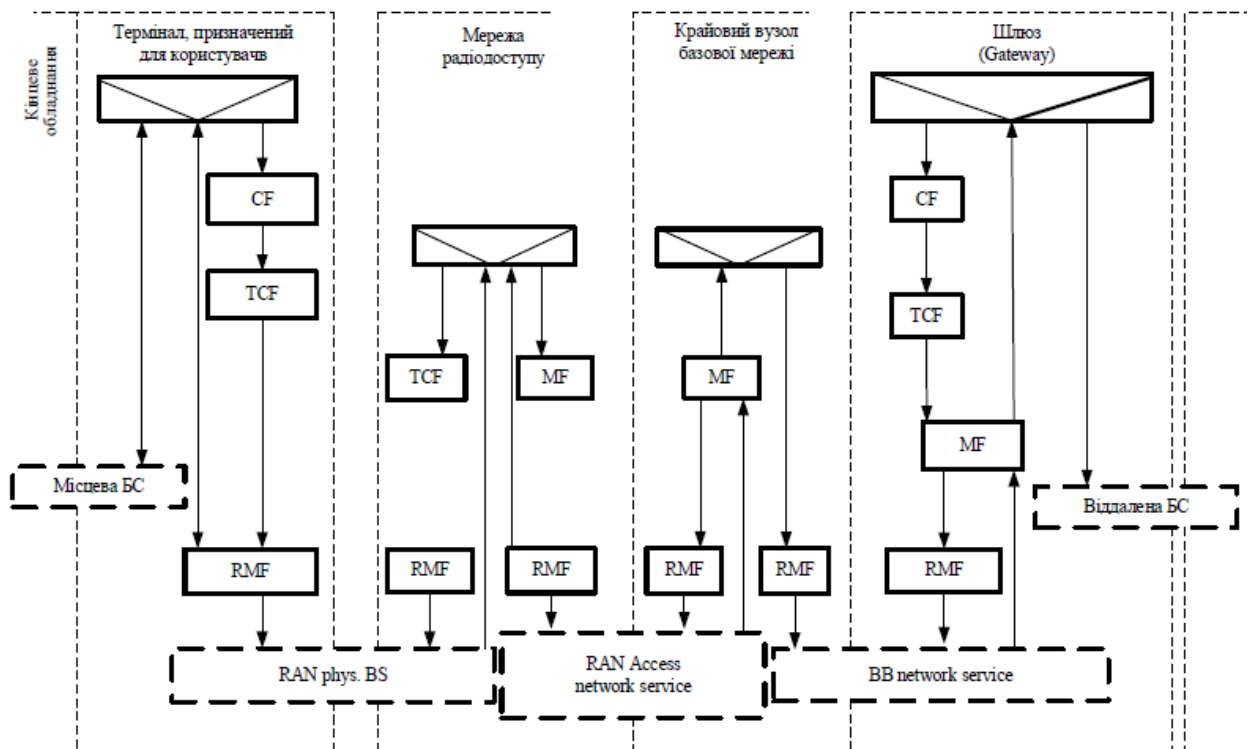


Рис. 2.7 Функції керування якістю обслуговування в призначеній для користувача площині

Основні функції мережі LTE, що належать до керування якістю обслуговування в площині керування

Функція управління послугами (SMF, Service Manager Function) є координуючою функцією при установці, модифікуванні і керуванні послугами, а також що координує функції керування якістю обслуговування в призначеній для користувача площині.

Функція (TF, Translation Function) *трансляції* перетворює внутрішні примітиви послуг мережі LTE в модулі різних протоколів взаємодіючих зовнішніх мереж, включаючи перетворення атрибутів послуг мережі LTE в параметри QoS протоколів зовнішніх мереж.

Функція управління можливостями (A/CCF, Admission / Capability Control Function) забезпечує інформацією про усі можливі ресурси мережевих вузлів, визначаючи при кожному запиті (чи модифікуванні) послуги, чи можуть мережеві вузли забезпечити необхідні ресурси. Ця функція також контролює можливість надання самої послуги, тобто чи реалізована в мережі запитана послуга.

Функція керування підпискою (SCF, Subscription Control Function) забезпечує контроль доступності абонентам певних послуг з необхідними параметрами QoS.

2.2.5. Реалізація концепції QoS у мережах LTE

Відповідно до рекомендації 3GPP TS 23.402 V15.1.0 (2017-09), модель QoS, яка застосовується в поєднанні з опорними точками на основі протоколу PMIP, не використовує ідентифікатори каналів у пакетах площини користувача. Вона базується на фільтрах пакетів та пов'язаних з ними параметрах QoS (QCI, ARP, MBR, GBR), що надаються системі доступу через позамаршрутну (off-path) сигналізацію.

PCRF сигналізує ті ж фільтри пакетів та пов'язані з ними параметри QoS через Gxa, Gxb та Gxc, що і через Gx; інакше кажучи, деталізація інформації QoS, яка передається через Gxa, Gxb та Gxc, така ж, як і через Gx.

Канал EPS з S5/S8 на основі протоколу PMIP і протоколу доступу E-UTRAN показано на Рис. 2.8.

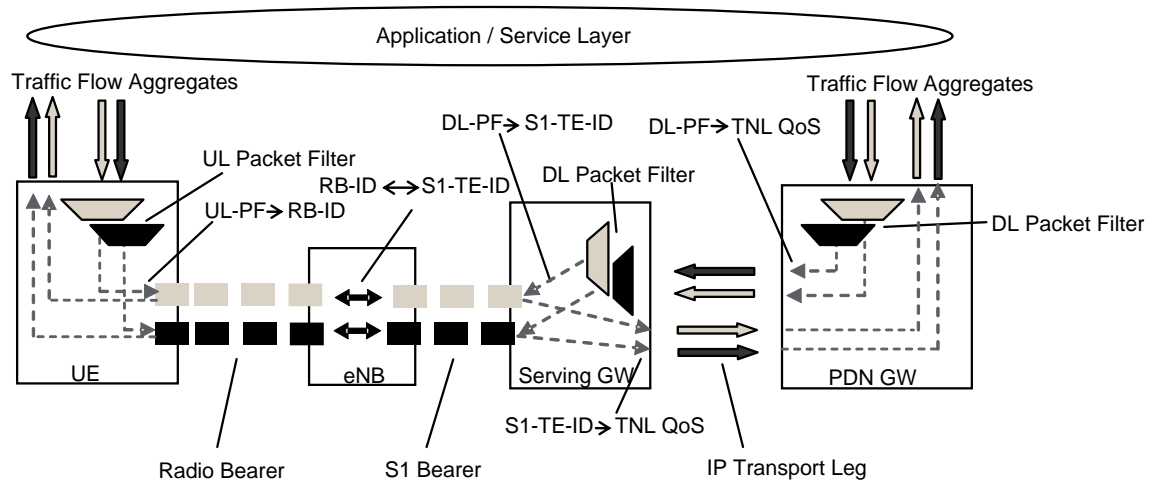


Рис. 2.8 Два канали Unicast EPS (S5/S8 на основі PMIP та протокол доступу E-UTRAN)

Для S5/S8 на основі протоколу PMIP та протоколу доступу E-UTRAN, канал EPS містить конкатенацію одного радіоканалу та одного каналу S1. Служба підключення PDN між обладнанням користувача (UE) і зовнішньою мережею пакетної передачі даних підтримується шляхом об'єднання каналу EPS та IP-з'єднання між шлюзом обслуговування і шлюзом PDN. Контроль QoS між Serving GW і PDN GW забезпечується на рівні транспортної мережі (TNL).

Канал EPS реалізується наступними елементами:

- В обладнанні користувача UFT TFT відображає агрегат потоку трафіка до каналу EPS у напрямку висхідної лінії зв'язку.
- У шлюзі обслуговування (Serving GW) DL TFT відображає агрегат потоку трафіку до каналу EPS у напрямку низхідної лінії зв'язку.
- радіоканал передає пакети каналу EPS між UE та eNodeB. Існує одноразове відображення між каналом EPS та радіоканалом.
- Канал S1 переносить пакети каналу EPS між eNodeB і шлюзом обслуговування. Існує одноразове відображення між каналом EPS та каналом S1.

- Для кожного обладнання користувача в туннелі PDN транспортуються пакети каналу EPS між шлюзом обслуговування та шлюзом PDN. Між каналом EPS та цим тунелем реалізується відображення «багато до одного».

- UE зберігає відображення між фільтром пакетної передачі висхідної лінії зв'язку та радіоканалом для створення зв'язків між агрегатом потоку трафіку та радіоканалом у висхідній лінії зв'язку.

- eNodeB зберігає відображення «один до одного» між радіоканалом та каналом S1, щоб створити зв'язок між радіоканалом та каналом S1 як у напрямку висхідної лінії зв'язку, так і в напрямку низхідної лінії зв'язку.

- Шлюз обслуговування зберігає відображення «один до одного» між фільтром пакетної передачі по низхідній лінії зв'язку та каналом S1, щоб створити відображення між агрегатом потоку трафіку та каналом S1 у низхідній лінії зв'язку.

- Шлюз доступу до інших мереж (PDN SW) забезпечує APN-AMBR для всіх SDF тих самих APN, які пов'язані з QCI без GBR.

2.2.6. Процедури виділеного каналу для протоколу доступу E-UTRAN з S5/S8 на основі PMIP

Процедура, зображена на рисунку Рис. 2.9, застосовується до всіх виділених операцій розподілу ресурсів для E-UTRAN, які активуються за допомогою PCRF, за винятком процедури дезактивації виділених каналів, ініційованої MME, яка розглядається в пункті 2.2.8. Процедури, ініційовані по S-GW в E-UTRAN відрізняються для кожного випадку.

Процедура, описана на Рис. 2.9, показує лише етапи, пов'язані з використанням S5/S8 на основі PMIP, які відрізняються від варіанту процедури GTP, наведеної в TS 23.401 [4].

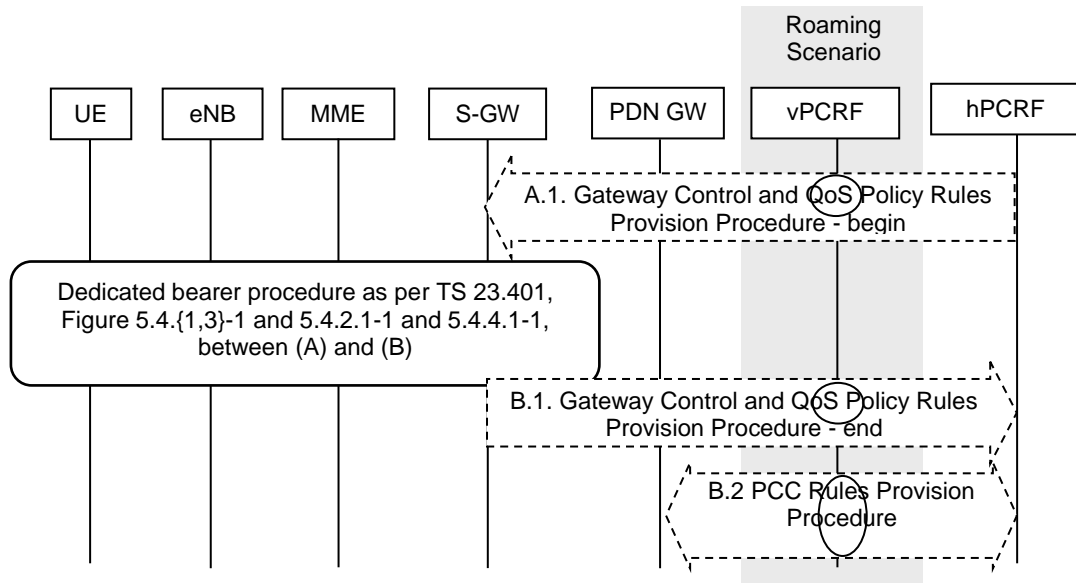


Рис. 2.9 Процедура розподілу ресурсів, UE в активному режимі

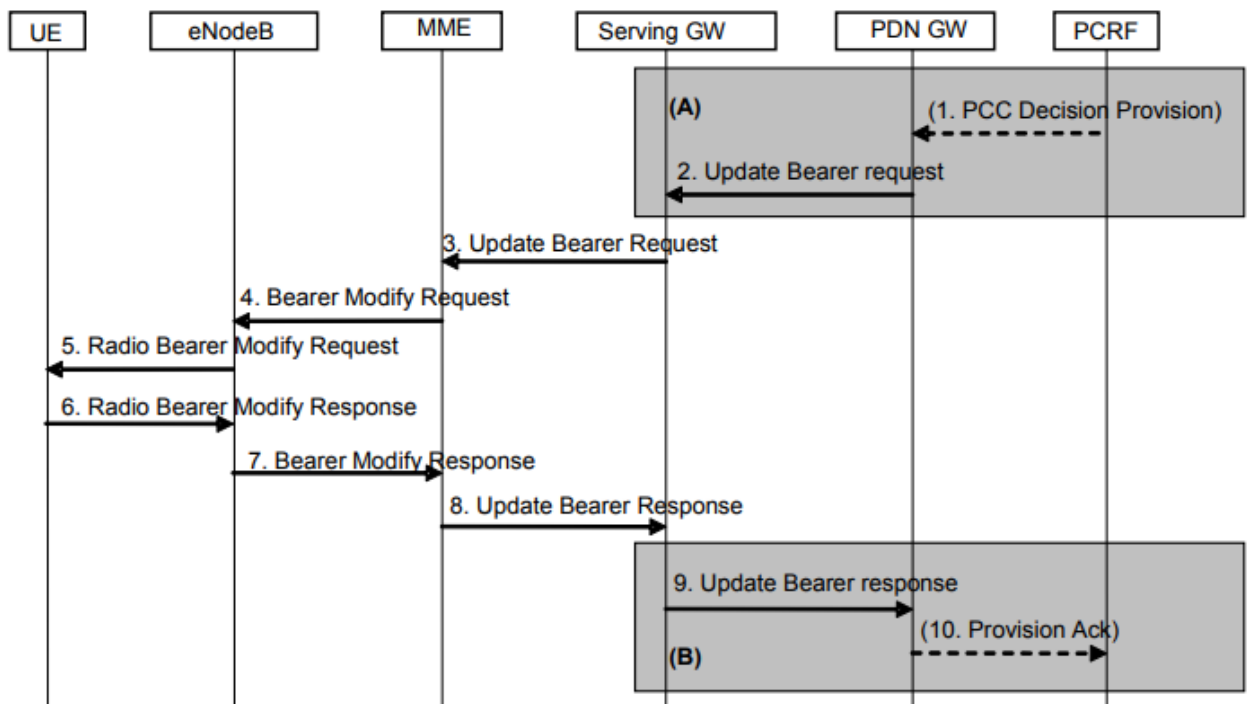


Рис. 2.10 Процедура модифікації носіїв з оновленим QoS носієм, UE в активному режимі

Ця процедура застосовується до випадків не роумінгу, роумінгу та місцевої обробки запитів (Local Breakout). Для випадків роумінгу та Local Breakout vPCRF пересилає повідомлення між шлюзом обслуговування і hPCRF. У разі Local Breakout vPCRF пересилає повідомлення між PDN GW та hPCRF.

Якщо динамічне положення політики не розгорнуто, кроки, наведені на малюнку, не виконуються. Замість цього може бути застосована налаштована статична політика.

A.1) PCRF ініціює процедуру постачання правил Gateway Control і QoS, зазначену в TS 23.203 [19], відправляючи повідомлення з правилами QoS та інформацією тригера подій до S-GW.

Кроки між A.1 і B.1 описані в TS 23.401 [4].

Для S5/S8 на основі PMIP, перед етапами процедури (крок 3 TS 23.401 [4], п. 5.4.1), PCRF надсилає повідомлення про прийняття рішення щодо керування політикою та тарікацією PCC (Policy and Charging Control) (політика QoS) до S-GW, а не до P-GW, як зроблено для S5/S8 на основі GTP. S-GW використовує цю QoS-політику, щоб визначити, що потоки трафіку повинні бути об'єднані або видалені з активного каналу. S-GW генерує TFT і оновлює якість обслуговування EPS-каналу, щоб відповідати агрегованому набору потоків трафіку. Функція зв'язування каналу S-GW може виконувати модифікацію, створення або видалення каналів у цій точці. Для модифікації S-GW надсилає MME повідомлення про оновлення запиту каналу (Update Bearer Request). Для створення виділеного каналу S-GW надсилає повідомлення "Створити канал" (Create Bearer), а для видалення, S-GW надсилає запит на видалення каналу (Delete Bearer Request).

B.1) Шлюз обслуговування вказує PCRF, чи може бути застосований запит на порядок правил політики QoS (QoS Policy Rules Provision), таким чином доповнюючи процедуру GW Control та QoS Rules Provision, розпочату на кроці A.1. Шлюз обслуговування повідомляє PCRF про місцезнаходження користувача та/або про часовий пояс UE (UE Time Zone) у якості звіту про події (Event Report), якщо встановлено відповідний тригер події.

B.2) PCRF ініціює процедуру надання правил PCC, як зазначено в TS 23.203 [19]. PCRF надає оновлені правила PCC для PCEF для виконання шляхом процедури PCC Rules Provision, зазначеної в TS 23.203 [19].

Крок B.2 може виконуватись перед кроком A.1 або паралельно з

кроками А.1 В.1, якщо для оновлення правил РСС в РCEF не вимагається підтвердження розподілу ресурсів. Детальніше описано в TS 23.203 [19].

2.3. Вплив віртуалізації мережевих функцій на процедури забезпечення якості обслуговування.

Незважаючи на успіхи, досягнуті в області розробки мереж мобільного зв'язку четвертого покоління, з'являються нові вимоги, викликані зростаючими потребами в комунікаціях, що в свою чергу вимагає розвитку нового покоління мобільних мереж (5G). Нові можливості використання, такі як потокове відео високої роздільної здатності, віддалений моніторинг, управління в реальному часі створюють вимоги, пов'язані з пропускнуою здатністю, затримкою передачі, надійністю і стійкістю мережі.

Очікується, що мережі забезпечать високу безпеку, мінімальну затримку (нижче 5 мс.), надійність передачі 99,999% і стовідсоткову доступність.

Розробка мережі 5G повинна бути спрямована на подолання вищевказаних обмежень, з метою надання ультра-надійних, безпечних сервісів з мінімальною затримкою великій кількості інтелектуальних об'єктів і систем, а також новим мобільних терміналів.

Порівняння технологій 3G/4G/5G представлено на Рис. 2.11. Технології IMT-2000 (3G) IMT-Advanced (4G) покращують пропускну здатність мережі, швидкість передачі даних користувачів, використання спектра і зменшують затримки. Впровадження технології IMT (5G) планується для "потужної і критично важливої машинної взаємодії", зменшення затримки, спектральної ефективності, швидкості, мобільності і надійності.

Мережева архітектура 5G складається з двох шарів: радіомережі та хмарної мережі. Різні типи базових станцій, що виконують мінімальний набір функцій, утворюють радіомережу. Хмарна мережа складається з площини користувача (User Plane Entity - UPE) і площини управління (Control Plane Entity - CPE), які виконують функції площини користувача і площини управління відповідно. Як показано на Рис. 2.12, фізична реалізація хмари

може бути адаптована для задоволення різних цільових показників. Наприклад, UPE і CPE можуть бути розташовані близько до базових станцій для зменшення затримки критичних послуг. Також може бути виконано підключення БС до невеликого прилеглого центру обробки даних (ЦОД-3), а не до центрального ЦОД-2. З іншого боку, БС може бути з'єднана з ЦОД-2, якщо затримка не критична. Така гнучкість дозволяє оператору розгорнути великі і малі центри обробки даних для підтримки конкретних потреб у послугах. Така архітектура спрощує мережу і забезпечує швидке і гнучке розгортання і управління.

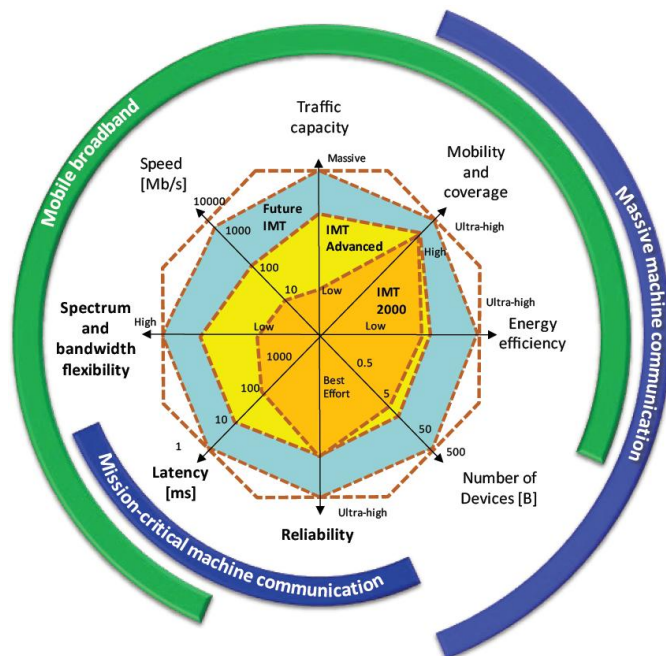


Рис. 2.11 Порівняння технологій 3G/4G/5G

Різні технології, такі як Network Functions Virtualization (NFV) і Software Defined Networking (SDN) призначені для створення і впровадження таких мереж. Проте, майбутні сервіси, такі як потокове відео та інші послуги мають різні вимоги, які підкреслюють необхідність динамічного масштабування функціональності мережі.

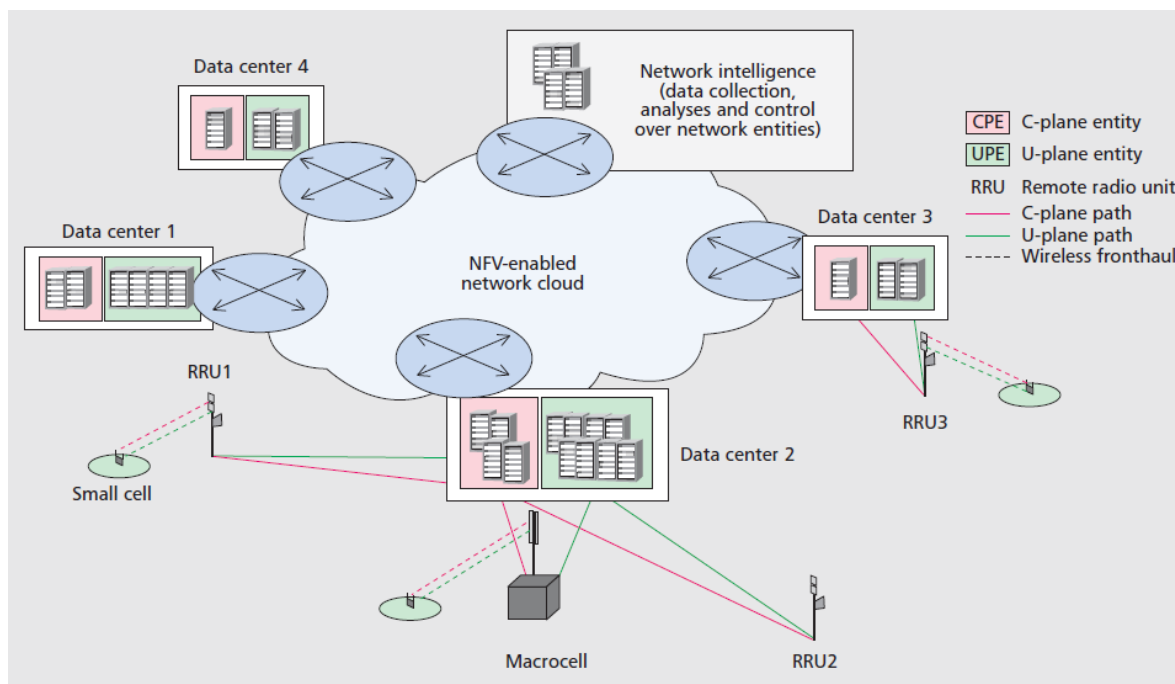


Рис. 2.12 Мережева архітектура 5G

Гетерогенні мережі відрізняються підвищеною пропускнуою здатністю, великим покриттям і надійністю, більш високою ефективністю використання радіочастотного спектру і низьким енергоспоживанням з боку як мережевої інфраструктури, так і термінального обладнання.

Мобільні оператори мережі стикаються з наростаючою проблемою зростання трафіку передачі даних, в зв'язку з поширеністю смартфонів і трансляванням аудіо і відео послуг. У новій парадигмі оператори повинні управляти навантаженням, задовольняючи зростаючі споживчі і корпоративні очікування продуктивності, забезпечуючи повсюдний широкосмуговий доступ, а також швидко впроваджувати нові послуги, щоб зберегти конкурентну перевагу. Існуючі мобільні мережі постійно стикаються з такими обмеженнями, як стаціонарне і дороге устаткування, складні протоколи управління і гетерогенні інтерфейси конфігурації. З метою вирішення поточних обмежень, необхідно вивчати і застосовувати принципи SDN в мобільних мережах, а саме SDMN (Software Define Mobile Networks). SDN розділяє рівні управління і передачі даних, використовуючи стандартні протоколи, що дозволяють віддаленим пристроям здійснювати управління і

експлуатацію рівнів даних. Протокол синхронізації потрібен для комунікації обох рівнів, одним з таких протоколів є OpenFlow [2]. Переваги SDN в області хмарних обчислень є очевидними, однак застосування даної концепції в мобільних мережах вимагає подальшого вивчення.

2.4. Модель керування енергозбереженням телекомунікаційної системи, навантаженням та ресурсами обслуговування

В умовах невідпинного розвитку інформатизації суспільства, постійного зростання одиниць електронних пристроїв та обладнання, що споживає електричну енергію виникла необхідність оптимізації систем керування процесом обслуговування в телекомунікаційних системах.

Зменшити споживання енергетичних ресурсів телекомунікаційною системою за рахунок керування навантаженням, процесом обслуговування в мережі радіо доступу, транспортній мережі та мережі ядра мобільного оператора, за рахунок впровадження комплексних моделей аналізу статистики, методів багатокритеріального оцінювання функціонування системи.

Телекомунікаційна мережа представлена у вигляді вузлів, вузлами телекомунікаційної мережі виступають різноманітні підсистеми, які мають різні технічні характеристики, а також виконують різноманітні функції такі як:

- Функції взаємодії з абонентськими терміналами (базові станції),
- Функції прийому та передачі інформаційних потоків у транспортній системі (комутаційне обладнання),
- Функції обслуговування службових інформаційних потоків (підсистеми ядра мобільного зв'язку).

Крім того вузли, що розглядаються також можна поділити на групи з подібними технічними характеристиками.

Метою роботи є зменшення енергетичних витрат на обслуговування телекомунікаційної мережі. На енергетичні витрати впливає низка факторів,

які відрізняються залежно від функцій обладнання, виробника, та програмно-апаратного забезпечення телекомунікаційного вузла.

Однак доведено, що енергоспоживання залежить від навантаження на систему, тобто від кількості операцій які виконує пристрій. Тому організація потоків навантаження на підсистеми телекомунікаційної мережі дозволить оптимізувати енергетичні витрати.

Суттєво зменшити витрати на обслуговування телекомунікаційної мережі доступу можна за рахунок зменшення потужності базових станцій. Зміна потужності базової станції призведе до низки пов'язаних з цим проблем. Однак на території міст, де покриття базовими станціями є дуже щільним, зменшення покриття однієї базової станції, що пов'язано зі зменшенням потужності суттєво не вплине на доступність послуг зв'язку, якщо параметри мережі базових станцій правильно розраховані.

Отже, постає задача: розрахувати параметрів для вузлів телекомунікаційної мережі таким чином, щоб загальне споживання електричної енергії було мінімальним, а якість надання послуг не змінилася.

Оскільки протягом доби навантаження на телекомунікаційну мережу змінюється, тому розрахунок та запровадження планів використання телекомунікаційного обладнання мобільного оператора дозволить зменшити витрати на енергоресурси, які забезпечують роботу телекомунікаційної мережі.

Постановка задачі:

N – кількість вузлів обслуговування, вузол обслуговування – група обладнання телекомунікаційної мережі, які обслуговують випадкове навантаження, що створює абонент, та може розглядатися як єдине ціле, та бути налаштоване відповідно вимогам моделі.

N_d – кількість вузлів обслуговування, що утворюють мережу доступу.

N_{tr} – кількість вузлів обслуговування, що утворюють транспортну мережу.

N_{core} – кількість вузлів обслуговування, що утворюють мережу ядра мобільного оператора.

N_1, N_2, \dots, N_M – підмножини вузлів обслуговування, на які розділяється множина N . Обладнання розташоване у вузлах, які входять в одну підмножину має однакові характеристики.

M – кількість можливих груп обладнання.

наприклад, N_1 – кількість базових станцій, що підтримують технологію GSM,

N_2 – кількість базових станцій, що підтримують технології 3G,

N_3 – кількість радіорелейних станцій,

N_4 – кількість SDH комутаторів

N_5 – кількість комутаторів DWDM, та інші.

$$N = \sum_{m=1}^M N_m$$

Крім того, в рамках однієї підмножини N_m ($m=1, \dots, M$) можна виділити обладнання від різних виробників, які можуть мати відмінні технічні характеристики та по різному споживати електричну енергію, відповідно потребують окремого налаштування. Тоді підмножина N_i розділяється на підмножини N_{ms} ($m=1, \dots, M, s=1, \dots, S$), де S – максимальна кількість різних типів обладнання, які входять до однієї підмножини.

Слід зауважити, що N – велике число, наприклад, кількість базових станцій одного оператора перевищує 7000, і це число постійно зростає. Так само зростає кількість комутаційного обладнання мережі.

Залежно від типу вузла телекомунікаційної мережі будуть застосовані різні засоби енергозбереження. Для вузлів мережі доступу в результаті оптимізаційної роботи буде запропоновано план зміни потужності випромінювання, та можливе відключення. А для транспортної мережі буде запропонована схема границь навантаження, обслуговування якого є енергоефективним.

G – матриця зв’язаності вузлів. Якщо $i \in N_d$ та $k \in N_d$, тоді g_{ik} – показує наявність сусідства базових станцій.

J – кількість типів сервісів, які створюють навантаження на вузли телекомунікаційної мережі. В даній математичній моделі ми обмежуємося трьома типами трафіку – голосовий трафік, трафік даних та службовий трафік, в разі необхідності можна деталізувати.

t – змінна часу, дискретна величина, яка показує момент часу в який було зроблено спостереження.

$\lambda_{ij}(t)$ – математичне очікування інтенсивності навантаження на вузол i ($i=1, \dots, N$) сервісом j -го типу ($j=1, \dots, J$) в момент часу t . Для розрахунку математичного очікування аналізується статистика для однорідних періодів часу. Наприклад, математичне очікування навантаження заданого типу сервісу в будній день о 10 годині. При описі математичної моделі ми не зупиняємось на аналізі статистики, розрахунку середніх значень навантаження, та групуванні періодів часу де навантаження є однорідним. Далі розглядається модель типового робочого тижня.

$\lambda_i(t) = \sum_j \lambda_{ij}(t)$ – математичне очікування сумарного навантаження на вузол i ($i=1, \dots, N$) в момент часу t .

Q_j – множина граничних значень показників якості обслуговування для j -го ($j=1, \dots, J$) типу сервісу. В результаті оптимізаційного процесу перерозподілу навантаження між вузлами телекомунікаційної мережі значення показників якості повинно зберегтися в межах заданих значень.

$q_{ij}(t)$ – значення показника якості q для j -го ($j=1, \dots, J$) типу сервісу у вузлі i ($i=1, \dots, N$) в момент часу t

$\mu_{ij}(t)$ – інтенсивність обслуговування сервісу j -го типу ($j=1, \dots, J$) у вузлі i ($i=1, \dots, N$) в момент часу t . Міркування відносно часу аналогічні до параметру $\lambda_{ij}(t)$. Як було доведено, зміна параметру сумарної інтенсивності обслуговування $\mu_i(t) = \sum_j \mu_{ij}(t)$ призводить до зміни енергоспоживання i -го вузла.

$E_i(t)$ – енергоспоживання i -го ($i=1,..N$) вузла в момент часу t .

$f_i(\mu_i)$ - функції залежності енергоспоживання i -го ($i=1,..N$) від інтенсивності обслуговування сумарного навантаження на вузол. Структура функцій залежності розраховується експериментально для кожної з M груп обладнання.

Аналіз функцій $f_i(\mu_i(t))$ показав, що можна виділити діапазон значень $\mu_i(t)$ для яких енергоспоживання E_i буде оптимальним.

$\lambda_{ij}'(t)$ – кількість запитів на обслуговування j -го типу ($j=1,..J$), які будуть надходити у вузол i ($i=1,..N$) в момент часу t після енергоефективного перерозподілу навантаження.

$\mu_{ij}'(t)$ – інтенсивність обслуговування сервісу j -го типу ($j=1,..J$) у вузлі i ($i=1,..N$) в момент часу t після енергоефективного перерозподілу навантаження.

Необхідно знайти розподіл навантаження $\lambda_{ij}'(t)$ та потужностей, що забезпечать інтенсивність обслуговування $\mu_{ij}'(t)$, при яких сумарне енергоспоживання телекомунікаційної мережі в момент часу t , $E(t) = \sum E_{i(t)}(t)$, буде мінімальним, а показники якості обслуговування сервісів $q_{ij}(t)$ будуть збережені у межах заданих множиною Q_j .

Розв'язок поставленої задачі є складною задачею, при розв'язанні якої я пропоную користуватися за наступними принципами:

1. Поетапно розв'язувати задачі оптимізації мережі доступу, транспортної мережі та для мережі ядра оператору зв'язку.
2. Врахувати джерела навантаження на вузли телекомунікаційної мережі.
3. Окремо розглядати вузли, для яких може бути застосовано спільний спосіб регулювання інтенсивності обслуговування.
4. Сформувані множини вузлів, між якими може бути перерозподілене навантаження.

Структура вхідного навантаження для мережі доступу. Навантаження на мережу доступу створюється мобільними пристроями, які зв'язуються з базовими станціями. Передача інформації від мобільного пристрою до базової станції можлива за умови наявності достатнього рівня сигналу від базової станції. Переключення між базовими станціями здійснюється за умови зниження рівня сигналу нижче допустимого значення.

Під зміною інтенсивності обслуговування для $i \in N_d$ мається на увазі зміна потужності випромінювачів, які забезпечують downlink передачу даних до мобільного пристрою від базової станції. При зменшенні потужності базової станції зменшується радіус її дії, також знижується рівень сигналу базової станції, який бачить мобільний пристрій.

Окремою задачею для вузлів $i \in N_d$ є *визначення нижньої границі потужності сигналу* відповідної базової станції, щоб покриття не змінилося.

Структура навантаження для транспортної мережі. Для вузлів $i \in N_{tr}$ вхідне навантаження має дві основні складові: навантаження, що надходить від вузлів мережі доступу $k \in N_d$, яке може бути налаштовано при рішенні оптимізаційної задачі для мережі доступу, та транзитне навантаження, яке надходить від інших вузлів транспортної мережі. При використанні SDN технології є можливість рівномірно розподіляти інформаційні потоки по транспортній мережі, так щоб забезпечити енергоефективність роботи транспортних вузлів.

Під новим значенням інтенсивностей обслуговування $\mu_{ij}'(t)$ для вузлів $i \in N_{tr}$ розуміється розрахунок верхньої границі об'єму трафіку j -го типу сервісу ($j=1, \dots, J$), що передається через вузол. При застосуванні технології програмно-керованих мереж можливе централізоване керування потоками у вузлах транспортної мережі. Відповідно об'єми трафіку, який проходить через транспортний вузол, можуть змінюватися залежно від часу доби, та будь-яких інших умов. Тобто енергозберігаюча схема транспортних потоків оператора зв'язку може бути застосована.

Енергоспоживання вузлів мережі ядра оператора зв'язку залежить від ефективності організації обчислювальних процесів. Обслуговування службових потоків у підсистемах ядра мобільного зв'язку, представляє собою послідовність стандартних операцій, виконання яких потребує обчислювальних ресурсів таких як оперативна пам'ять, постійна пам'ять, процесорний час, та ресурс внутрішньої мережі. Обслуговування службових потоків здійснюється на декількох обслуговуючих одиницях (серверах), які розміщені територіально. На сьогоднішній день системи обслуговування службового трафіку тісно пов'язані з спеціальним обладнанням, де створюються службові потоки, та спеціальним (вендорним) програмним забезпеченням, де вони обслуговуються. Однак поступовий перехід до віртуалізації окремих функцій та систем дозволить гнучке керування системою обслуговування службовими потоками. Навантаження на вузли $i \in N_{core}$ створюється при обслуговуванні службового трафіку. Проблеми обслуговування службового трафіку, оптимізація структури ядра оператора зв'язку буде розглянута нижче.

У дисертаційній роботі запропоновано єдиний підхід до вирішення проблеми енергозбереження у вузлах телекомунікаційної мережі, який полягає в тому, щоб розподілити навантаження між доступними вузлами обслуговування, щоб забезпечити завантаження системи на заданому рівні, який забезпечує енергоефективне функціонування системи, крім того на заданому рівні зберігається показники якості обслуговування.

За основні прототипи математичної моделі, що зв'язує інтенсивність надходження запитів у вузол обслуговування телекомунікаційної мережі, інтенсивність обслуговування та показники якості функціонування обрано моделі К. Жернового. У статтях К. Жернового розглядаються системи масового обслуговування, де швидкість обслуговування пакету трафіку залежить від кількості заявок у черзі. Окремо розглядаються випадки з обмеженою та необмеженою чергою. Запропоновано розрахунок стаціонарного розподілу ймовірностей перебування системи у можливих

станах, де під станом системи розуміється одночасне перебування в системі заданої кількості пакетів. В даній системі параметром керування виступає швидкість обслуговування пакету в залежності від кількості заявок у черзі. В результаті можна підібрати такий план швидкостей обслуговування, для яких результуючий розподіл буде задовольняти вимогам до показників якості обслуговування трафіку. Також є формули для розрахунку середньої зайнятості системи обслуговування.

Висновки

1. Запропонована методологія забезпечення показників якості обслуговування гібридного сервісу у гетерогенному телекомунікаційному середовищі базується на чотирьох основних підходах: формування навантаження на вузол обслуговування, вибір потужності вузла обслуговування, визначення порядку роботи вузлів обслуговування та поточний контроль ефективності роботи системи, що дозволило забезпечити ефективне функціонування гетерогенного середовища обслуговування гібридних телекомунікаційних сервісів, а саме зменшити відсоток запитів, що обслуговуються в системі більше допустимого часу, визначеного експертами, - на 2%, зменшити відсоток запитів, які було втрачено через перевищення допустимого часу обслуговування на 3%, зменшити відсоток часу коли ресурси завантажені: менше за допустиме значення - на 8% та більше за допустиме значення на 10%.

2. Запропоновано модель керування інфраструктурою NFV, яка враховує особливості розташування дата центрів обслуговування, а також топологічної структури потоків гібридних телекомунікаційних сервісів, які поступають на обслуговування у дата центри, дозволяє забезпечити гнучке керування інформаційно-телекомунікаційною системою організованою із застосуванням хмарних обчислень.

РОЗДІЛ 3

СИСТЕМА КЕРУВАННЯ ГЕТЕРОГЕННОЮ ТЕЛЕКОМУНІКАЦІЙНОЮ СИСТЕМОЮ

Задачі керування телекомунікаційними системами активно досліджуються протігом часу їх існування. В класичній теорії керування [25-27] телекомунікаційнимим системами, визначено основні функції та підходи системи керування.

З розвитком концепції NGN функції керування відокремлюють від функцій передачі даних, розробляють моножину підходів та концепцій щодо функціонування системи керування телекомунікаціями.

Система керування телекомунікаціями. На сьогоднішній день активно досліджують та розвивають підходи до побудови програмно-керованих мереж, робота яких передбачає створення системи керування телекомунікаційною системою з використанням програмних контролерів, які складаються з компонентів

3.1. Цілі і завдання керування на телекомунікаційних мережах

Телекомунікаційні мережі являють собою досить складну організаційно-технічну систему, яка функціонує в умовах змінних внутрішніх і зовнішніх обставин. Ці зміни можуть виникати в результаті наступних причин:

1. Виходу з ладу або введення в експлуатацію окремих гілок, КЦ або цілих ділянок мережі.
2. Різкого збільшення величини навантаження в одному або декількох напрямках мережі зв'язку.
3. Зміна місцеположення абонентів або елементів телекомунікаційної мережі.
4. Підключення нових джерел навантаження та багато інших факторів.

Вплив цих факторів на ТКМ може призвести до:

1. Зміни складу шляхів передачі інформації в напрямках зв'язку і їх значимості.
2. Нестачі ресурсу мережі для передачі необхідних об'ємів інформації.
3. Виникнення місцевих чи загальномережевих перенавантажень.
4. Погіршення якості обслуговування чи повної відсутності зв'язку в деяких напрямках.

Це призводить до необхідності постійного контролю функціонування ТКМ і внесення відповідних корективів як в побудову мережі, так і в процеси, які в ній відбуваються. Вирішення цих завдань передбачає організацію і впровадження вискоєфективного керування. Його мета полягає в забезпеченні своєчасного встановлення з'єднань за заданими адресами і передачі визначеного об'єму повідомлень з дотриманням заданих імовірісно-часових характеристик при мінімальних матеріально-технічних затратах.

Поставлена мета може бути реалізована шляхом вирішення наступних задач:

1. Керування встановленням з'єднань за заданими адресами у відповідності з прийнятою системою обслуговування.
2. Підтримка показників основних характеристик ТКМ в межах необхідних норм при мінімальних матеріально-технічних затратах.

Перша задача визначається як задача керування встановленням з'єднань, друга – як керування ТКМ.

Керування встановленням з'єднань – це сукупність операцій з вибору і зайняття вільного шляху передачі інформації, відповідно до вказаної в запиті адреси. *Метою керування встановленням з'єднань* є забезпечення вибору оптимального шляху передачі даних, який враховує поточні параметри мережі зв'язку при обслуговуванні кожної заявки.

Керування телекомунікаційною мережею – це сукупність операцій, які забезпечують підтримку основних імовірісно-часових характеристик в межах потрібних норм. *Метою керування ТКС* є забезпечення показників

основних характеристик мережі, близьких до оптимальних значень, а у випадку їх відхилення – визначення шляхів доведення до необхідних норм при мінімальних матеріально-технічних затратах.

3.2. Системи керування на телекомунікаційних мережах

Враховуючи відмінності в цілях і характері процесів, які протікають в ТКМ, для забезпечення режиму нормального функціонування мережі зв'язку створюється дві системи: система керування встановленням з'єднань (СКВЗ) і система керування телекомунікаційною мережею (СКТМ).

Головними функціями СКТМ є:

- формування плану розподілу навантаження (ПРН), який забезпечує оптимальний режим функціонування мережі за одним чи декількома показниками;
- збір і обробка даних про стан елементів мережі;
- своєчасне визначення відхилень характеристик мережі від необхідних норм;
- визначення джерел дестабілізації мережі і шляхів доведення її характеристик до необхідних норм;
- прийняття рішень про здійснення відповідних корективів на ТКМ;
- визначення конкретних об'єктів керування;
- формування команд керування і передача їх у виконавчі органи.

Об'єктами керування СКТМ є:

- системи керування встановленням з'єднань;
- кросові пристрої елементів мережі зв'язку;
- сили і засоби, які забезпечують резервування, побудову і експлуатацію ТКМ.

Основними функціями СКВЗ є:

- прийняття запитів на встановлення з'єднань в необхідному напрямку зв'язку;

- визначення дисципліни обслуговування запитів, що надходять до мережі;
- визначення шляхів передачі даних, відповідно до адреси, вказаної в запиті;
- вибір шляху передачі повідомлення з урахуванням стану елементів мережі в момент надходження заявки;
- зберігання плану розподілу навантаження і його коректування за командами СКТМ;
- збір даних про результати обслуговування заявок;
- формування команд керування на виконавчі органи;
- передача необхідної інформації за запитами СКТМ.

Об'єктами керування СКВЗ є комутаційні прилади і прилади з обслуговування КЦ.

3.2.1. Узагальнена модель процесу керування телекомунікаційною мережею

Аналіз процесів, які протікають під час керування мережами зв'язку, дозволяє виділити чотири основні компоненти: ТКМ, систему керування телекомунікаційною мережею, систему керування встановленням з'єднань та фактори, що на них впливають.

ТКМ є об'єктом керування. Вона являє собою сукупність функціонально пов'язаних елементів, які взаємодіють між собою і абонентами.

СКТМ являє собою сукупність органів керування. СКВЗ виступає у двоякому вигляді. З одного боку, для мережі зв'язку це сукупність органів керування. З іншого боку, для СКТМ вона виступає в ролі об'єкта керування.

В процесі функціонування, як на органи, так і на об'єкти керування впливають різні фактори.

В даній моделі керування телекомунікаційною мережею можна виділити два вагомих аспекти. Перший визначається наявністю постійного

інформаційного обміну між її елементами. Фактично будь-яка завершена дія в елементах моделі тягне за собою формування і передачу повідомлення. Другий аспект проявляється у впливі на елементи моделі різних факторів, зумовлених як внутрішніми, так і зовнішніми процесами. Перші протікають в елементах мережі СКВЗ і СКТМ, другі – в навколишньому середовищі. Ці фактори багато в чому визначають режим інформаційного обміну між органами і об'єктами керування.

Фактори, які визначають характер функціонування системи керування, можна розділити на три групи:

1. Пов'язані з плановими структурними змінами в керованій мережі ($n_1, n_2, \dots, n_i, \dots, \in N$; $m_1, m_2, \dots, m_i, \dots \in M$) – розгортання і згортання елементів мережі, введення в експлуатацію і переміщення КЦ, зміна структури мережі і т.д.

2. Обумовлені виходом із ладу елементів керованої мережі в результаті агресивного впливу зовнішнього середовища і технічних відмов ($\psi_1, \psi_2, \dots, \psi_i, \dots, \in \Psi$) – ураження елементів чи ділянок мережі, виникнення несправностей в засобах зв'язку.

3. Визначені потребами абонентів в інформаційному обміні ($\lambda_1, \lambda_2, \dots, \lambda_i, \dots, \in A$) – зміни інтенсивностей інформаційних потоків і часу обслуговування заявок.

В процесі функціонування мережі, між її елементами і СКТМ здійснюється інформаційний обмін, який характеризується наявністю різного типу даних. Так від об'єктів керування до органів керування може передаватись наступна інформація:

- про відновлення режиму нормального функціонування елементів мережі ($h_1, h_2, \dots, h_i, \dots, \in H$);
- про ураження елементів телекомунікаційної мережі і технічні відмови, які ведуть до погіршення якості функціонування напрямків зв'язку (j_1, j_2, \dots ,

$j_i, \dots, \in J$) – вихід з ладу елементів чи ділянок мережі, систем передачі, лінійного і станційного обладнання КЦ, експлуатаційні помилки и т.д.;

- про зниження показників якості обслуговування абонентів різних категорій ($q_1, q_2, \dots, q_i, \dots, \in Q$) – перевищення величини втрат в напрямках мережі нормованих значень, перевищення часу очікування обслуговування.

Від органів керування на об'єкти керування можуть передаватись наступні команди:

1. На систему керування встановленням з'єднань:

- про необхідність зміни порядку вибору шляхів встановлення з'єднання ($u_1, u_2, \dots, u_i, \dots, \in U$) в напрямках зв'язку;
- про необхідність введення чи відміни визначених додаткових видів обслуговування ($d_1, d_2, \dots, d_i, \dots, \in D$).

2. На елементи телекомунікаційної мережі:

- про необхідність видачі результатів контролю технічного стану ($s_1, s_2, \dots, s_i, \dots, \in S$) елементів телекомунікаційної мережі;
- на введення в експлуатацію необхідних сил і засобів ($v_1, v_2, \dots, v_i, \dots, \in V$) для відновлення пошкоджених елементів чи ділянок мережі.

3. На керуючі пристрої КЦ:

- для здійснення кросових перемикань на КЦ, які забезпечують маневр каналами і трактами, для підтримки заданої пропускну здатності чи живучості напрямків зв'язку керованої мережі ($b_1, b_2, \dots, b_i, \dots, \in B$);
- про необхідність корекції програми роботи керуючих пристроїв КЦ ($z_1, z_2, \dots, z_i, \dots, \in Z$).

Дана модель описує процес керування телекомунікаційною мережею на досить високому рівні загальності. Тому при аналізі і синтезі конкретних систем керування необхідна деталізація особливостей їх побудови і методів вирішення задач керування, а також використовуваних протоколів взаємодії.

3.2.2. Класифікація систем керування телекомунікаційними мережами

Наявність особливостей в цілях і задачах керування на ТКМ, а також відмінності в способах вирішення цих задач, алгоритмах функціонування і вимогах, зумовлюють наявність декількох типів систем керування ТКМ. Тому необхідна класифікація СКТМ за їх відмінними ознаками.

З точки зору взаємодії з мережею зв'язку їх можна розділити на *статичні* і *динамічні*.

Під *статичними* розуміють СКТМ, які вирішують задачі керування. При цьому в якості вихідних даних використовуються прогнозовані значення параметрів мережі.

Вони діляться на:

1. Системи з незмінним ПРН.
2. Системи з ПРН, який змінюється за розкладом.
3. Системи з ПРН, який змінюється за ситуацією.

Перші вирішують задачі планування зв'язку при проектуванні ТКМ. В процесі функціонування таких СКТМ за заданими вихідними даними синтезується структура і топологія мережі, формується ПРН, визначаються потужності гілок, які забезпечують задану якість обслуговування в напрямках зв'язку.

Другі і треті вирішують задачу формування ПРН в переломленні до деяких прогнозованих стандартних ситуацій на ТКМ. При цьому, в других СКТМ зміна одного ПРН на інший здійснюється чітко за розкладом. В третіх – за командою оператора в момент виникнення стандартної ситуації.

Динамічними називаються СКТМ, які забезпечують керування реально функціонуючими мережами зв'язку за принципом «контроль – вплив – контроль». Їх основна задача полягає у контролі над параметрами функціонування мережі і прийняття рішень з необхідних коректувань у випадку їх відхилення від встановлених норм.

З точки зору можливостей і виду вирішуваних задач керування, динамічні СКТМ діляться на *оптимізаційні* та *ситуаційні*. Перші забезпечують вирішення всіх задач керування в повному обсязі, включаючи визначення джерел дестабілізації функціонування мережі і методів доведення їх характеристик до необхідних норм.

Можливості других обмежені. В основному, вони дозволяють підвищити якість функціонування мереж зв'язку за одним чи кількома параметрами. Але забезпечити підтримку характеристик мережі в межах заданих значень в умовах впливу дестабілізуючих факторів вони не можуть.

За швидкістю реагування за зміни в мережі, СКТМ можна поділити на *оперативні, статистичні і комбіновані*. *Оперативні* здійснюють негайне реагування на виникнення аварійних ситуацій в телекомунікаційній мережі. В основному дані СКТМ реагують на структурні зміни мережі зв'язку: вихід з ладу чи введення в експлуатацію гілок мережі, аварії чи переміщення КЦ і т.д. За сигналами про зміни в структурі мережі зв'язку оперативні СКТМ виконують переформування ПРН і доведення його до виконавчих органів.

Статистичні системи призначені для виявлення змін ймовірно-часових характеристик ТКМ. Це можуть бути зміни інтенсивностей потоків заявок, величини втрат на гілках і напрямках зв'язку і т.д. Реакція цих систем на зміни ситуації в мережі зв'язку значно нижче, ніж оперативних. Це пов'язано з необхідністю накопичення певного об'єму статистичних даних про процеси, які відбуваються в мережі, що дозволяє робити висновки з певною точністю і достовірністю. І тільки після цього, за результатами аналізу змін, виконувати корекцію плану розподілу навантаження.

Комбіновані системи передбачають спільне використання методів керування статистичних і оперативних СКТМ. В процесі вирішення задач керування вони формують ПРН, який враховує не лише стан структури мережі, але і ймовірно-часові параметри інформаційних потоків.

Класифікація систем керування мережами представлена на Рис. 3.1.

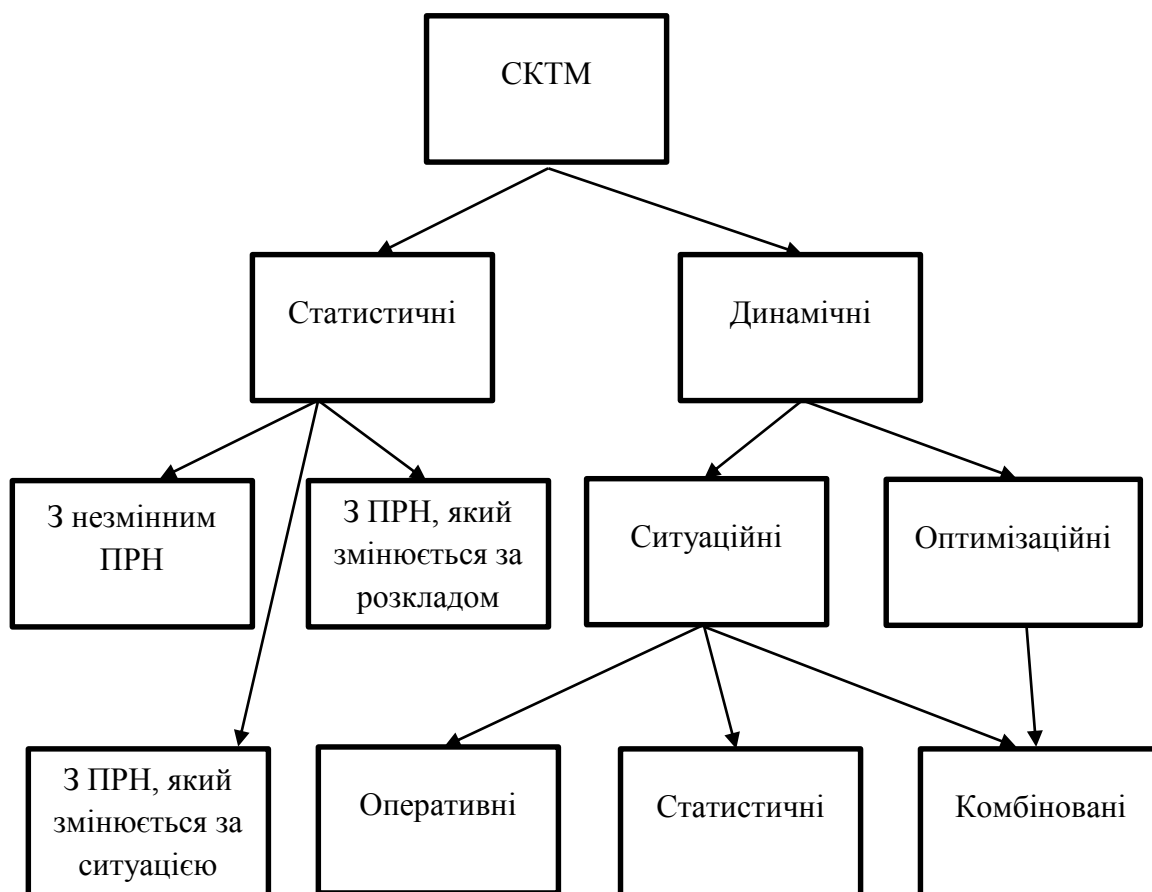


Рис. 3.1 Класифікація систем керування мережами

3.2.3. Функції організаційних систем керування

В залежності від особливостей реалізації систем керування телекомунікаційними мережами, в сучасній науково-технічній літературі виділяють централізовані, зонові, децентралізовані і комбіновані принципи структурної побудови СКТМ.

Централізовані системи керування телекомунікаційними мережами

Централізовані СКТМ передбачають наявність центрального органу керування мережею (ЦОКМ) і місцевих виконавчих органів (МВО), які розташовуються, як правило, спільно з КЦ мережі і здійснюють як довготривалу, так і оперативну комутацію.

Основними функціями центрального органу керування є:

- формування плану розподілу навантаження телекомунікаційної мережі;
- збір і обробка інформації про стан елементів навантаження елементів мережі;
- оцінка завантаженості ділянок мережі;
- виявлення відхилень параметрів потоків заявок від нормованих значень;
- визначення можливостей перерозподілу навантаження з метою підвищення пропускної здатності мережі зв'язку чи окремих її ділянок;
- коректування плану розподілу навантаження;
- формування команд для МВО.

Місцеві виконавчі органи виконують наступні основні функції:

- контроль технічного стану елементів керованої ділянки мережі;
- збір даних про параметри потоків інформації в напрямках зв'язку;
- передача результатів контролю на ЦОКМ;
- прийом команд від ЦОКМ;
- дія на об'єкти керування СКТМ (криси, програми СКВЗ, сили і засоби, які виконують побудову магістралей).

Місцеві виконавчі органи розташовуються, як правило, на КЦ. Центральний орган керування може розташовуватись як на одному з КЦ телекомунікаційної мережі, так і в окремому спорудженні.

Зонові системи керування телекомунікаційними мережами

Побудова зонових СКТМ передбачає розбиття телекомунікаційної мережі на ділянки, які називаються зонами. Основні вимоги до зони мережі зв'язку:

- територіальна обмеженість;
- зручність керування елементами зони;
- єдина підпорядкованість сил і засобів, які забезпечують побудову, експлуатацію і технічне забезпечення засобів зв'язку зони.

В кожній зоні керування здійснюється централізовано зоновим органом керування (ЗОКМ). Периферійним обладнанням СКТМ є МВО.

Децентралізовані системи керування телекомунікаційними мережами

При децентралізованій побудові систем керування мають місце єдині місцеві органи керування (МОКМ). Вони розміщуються, як правило, спільно з КЦ. Основними функціями МОКМ є:

- формування і зберігання ПРН;
- контроль технічного стану вихідних від КЦ гілок;
- збір інформації про імовірісно-часові характеристики інформаційних потоків в напрямках зв'язку;
- оцінка завантаженості вихідних від даного КЦ гілок потоками вимог різних напрямків;
- виявлення відхилень якості обслуговування заявок по гілках і напрямках зв'язку від нормованих значень;
- визначення можливостей оптимізації функціонування керованої ділянки телекомунікаційної мережі;
- внесення змін в ПРН;
- формування команд керування і доведення їх до виконавчих органів, які відносяться до керованої ділянки мережі.

Комбіновані системи керування телекомунікаційними мережами

Комбіновані СКТМ мають в своєму складі як елементи централізації, так і децентралізації керування. При цьому МОКМ на основі наявної інформації вирішують окремі задачі в інтересах свого КЦ, а центральний орган оптимізує функціонування всієї мережі в цілому, вносячи корективи в роботу місцевих органів.

З точки зору якості керування, найбільші можливості мають централізовані системи, які окрім формування плану розподілу навантаження за критерієм мінімальної довжини шляху, дозволяють приймати оптимальні рішення щодо методів доведення характеристик мережі до необхідних норм.

Можливість вирішення основних задач керування мережею зв'язку в повному об'ємі дозволяє вважати децентралізований спосіб побудови СКТМ перспективним, але його низька живучість обмежує можливості практичного використання.

3.3. Програмно визначені мережі

В традиційно використовуваних комп'ютерних мережевих пристроях, таких як маршрутизатори та комутатори, типова конфігурація полягає у тому, рівень керування і рівень передачі даних поєднані один з одним. І навіть якщо вони роз'єднані, обидва функціонують на одному і тому ж пристрої і потребують ре конфігурації. Рівень передачі даних відповідає за передачу вхідних пакетів через вихідні порти на основі існуючих правил, в той час як рівень керування приймає рішення про те, яку дію потрібно виконати над кожним пакетом шляхом встановлення відповідних правил передачі даних.

Програмно визначена мережа (SDN) це сучасна комп'ютерна мережева архітектура, яка дозволяє фізичний розподіл рівня керування та рівня передачі даних мережі. Дана концепція надає можливість керувати кількома платформами передачі даних через рівень керування, який контролюється єдиною сутністю, дозволяючи абстрагувати існуючу мережеву архітектуру з точки зору прикладних програмних компонентів. Ця абстракція дозволяє користувачеві керувати мережею з глобальної точки зору, замість того, щоб контролювати окремо кожен пристрій.

За допомогою SDN рівень керування може керуватись за допомогою програмного забезпечення – контролера. Він зв'язується з фізичними і віртуальними комутаторами через загальний протокол, незалежно від їх виробників.

Найбільш використовуваним протоколом SDN є OpenFlow (OF), незважаючи на те, що існує ряд інших відомих протоколів конфігурації пристроїв, таких як SNMP (Simple Network Management Protocol) або

NETCONF (Network Configuration). Цей протокол та відповідні інтерфейси, які використовують для встановлення з'єднання між контролером та пристроєм, що підтримує OpenFlow, є частиною інтерфейсу контролера південного напрямку (Southbound), в той час як інтерфейс північного напрямку (Northbound) дозволяє зовнішнім прикладним програмним компонентам взаємодіяти з контролером через спеціальні інтерфейси – Application Programming Interfaces (APIs). Інтерфейс південного напрямку відрізняється лише використанням протоколом та сумісністю версій на мережевих пристроях та контролері, а інтерфейс північного напрямку є незалежним і специфічним для кожного контролера. Крім того, існує також поняття інтерфейсів східного і західного напрямків, які дозволяють взаємодію кількох контролерів. Прикладом цієї взаємодії може бути *Гіперпотік* – розподілений на основі подій рівень керування для OpenFlow, в якому декілька контролерів можуть обмінюватись даними і ділитись глобальним виглядом мережі між собою. На рис. 3.2 зображено існуючий інтерфейс і компоненти архітектури SDN.

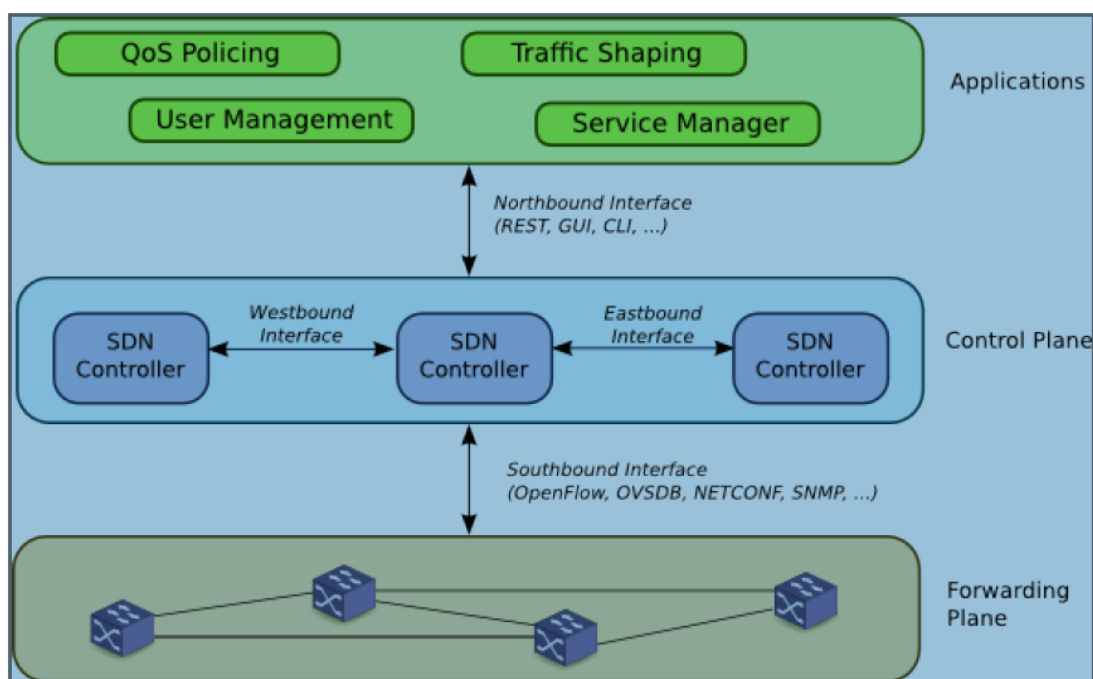


Рис. 3.2 Діаграма архітектури програмно-визначеної мережі

3.3.1. Передумови впровадження SDN

Так як SDN є досить новою мережевою концепцією, завжди існують певні протидії щодо перспектив переходу від існуючих мереж, які включають такі аспекти, як продуктивність, гнучкість, масштабованість, безпеку та сумісність. Це одні з основних викликів, які визначають подальший розвиток в даному напрямку. Крім того, впровадження даної концепції вже можна спостерігати в мережах великих центрах обробки даних, таких як Google.

В той час як продуктивність мережі SDN безпосередньо пов'язана з продуктивністю використовуваних пристроїв (яку визначає затримка та пропускна здатність), гнучкість відноситься до можливості впровадження нових функцій в системі. Масштабованість орієнтується на поведінку контролера та його можливість взаємодіяти з іншими контролерами і множинними мережевими вузлами.

Так як в майбутньому планується широке використання концепції SDN, безпека є досить важливим аспектом (наприклад, зловмисники можуть маскуватись під контролер і здійснювати атаки на мережеві комутатори). FortNox – є розширенням програмного забезпечення SDN контролера яке реалізує авторизацію в системі SDN на основі ролей. З рольовою авторизацією захищені прикладні програмні компоненти ідентифікуються, тому встановлення конфліктних правил потоку з невідомими прикладними програмними компонентами може бути виявлене, і відповідні потоки будуть відхилені.

Реалізація продуктів SDN в існуючих мережах безпосередньо пов'язана з ключовою проблемою сумісності. Заміна усіх мережевих пристроїв на ті, що підтримують SDN, може бути складним рішенням через адміністративні та фінансові витрати. З іншого боку, розширення складності задач, які виконують оператори мережі, є досить помітною при реалізації рішень SDN в існуючій мережі, так як вона збільшує кількість одночасно виконуваних задач і, в той же час, зберігає зв'язність існуючих мережевих пристроїв.

Таким чином, одночасне використання пристроїв існуючої мережі та тих, що підтримують SDN, вимагає спеціальних керованих рішень.

Приклади поточного використання програмно визначених мереж можна знайти в літературі. Вони можуть бути пов'язані зі створенням і впровадженням сервісів оцінки якості обслуговування, які реалізуються на контролерах OpenFlow чи інших мережах з підтримкою SDN. Для передачі відео, голосу, онлайн ігор та інших прикладних програмних компонентів, які працюють в реальному часі, досягнення необхідних показників якості обслуговування є однією з найважливіших задач. QoS можуть пропонувати провайдери послуг, надаючи відповідні прийнятні значення для різних мережних метрик, такі як гарантована пропускна здатність, затримка end-to-end, джиттер та допустима кількість втрачених пакетів. При використанні концепції SDN, наявність цілісної картини мережі може бути великою перевагою при конфігуруванні комутаторів, так як даний вигляд мережевої топології надає вхідні дані для розрахунку QoS-залежних потоків.

Розширення SDN контролера по протоколу API північного напрямку надає додаткову підтримку для онлайн інтерактивних програм реального часу (ROIA – Real-Time Online Interactive Applications), дозволяючи їм запитувати і вказувати відповідні параметри QoS і вимоги мережі.

Однією з переваг, які надає OpenFlow є прикладний програмний компонент *MonSamp*. Це програма моніторингу, яка збирає метрики QoS від існуючих потоків в SDN комутаторах. Іншою платформою для моніторингу мережі є OpenSample. Вона значно зменшує інтервал випробування порту і підрахунок статистики потоку шляхом використання інструменту моніторингу трафіку sFlow, який підтримується на багатьох моделях комутаторів в якості альтернативи стандартним можливостям OpenFlow. Для виявлення аномалій мережі може використовуватись інше рішення sFlow на основі методу ентропії.

3.3.2. SDN контролери

По мірі зростання популярності програмно визначених мереж кількість доступних SDN контролерів також зростає. Під час вибору SDN контролера потрібно враховувати і порівнювати декілька аспектів, таких як протоколи API (північного, південного, східного і західного напрямків), базові мережеві сервіси, доступні інтерфейси керування, базова архітектура, використовувана мова програмування, розмір і досяжність спільноти користувачів та розробників та підтримка версій OF.

Як уже згадувалось раніше, API інтерфейси південного напрямку використовують для забезпечення взаємозв'язку між рівнями керування та передачі даних існуючих пристроїв. В той час як протокол OF є основним протоколом, який використовується в цій частині архітектури SDN, реалізацію інших відомих протоколів можна знайти в SDN контролерах. Прикладами цих протоколів, які використовують в якості альтернативи до OpenFlow для встановлення з'єднання між рівнем керування і передачі даних пристроїв, є протоколи OpFlex, POF (Protocol-oblivious forwarding), ForCES (Forwarding and control element separation).

Інші протоколи можуть використовуватись в якості доповнень до пристроїв керування конфігурацією. Це такі протоколи, як OVSDB (Open vSwitch Database Protocol) – використовується для взаємодії з Open vSwitch (OVS) та керування його конфігурованими характеристиками. OVS – це набір програмованих комутаторів, який окрім підтримки правил потоку OF містить додаткові характеристики (такі як конфігурація QoS, тунелювання протоколу IP, керування несправностями підключення та підтримка мережі VLAN). Також використовують протоколи SNMP та NETCONF для конфігурування та взаємодії старих мережевих пристроїв.

Для контролерів, які підтримують OpenFlow в якості протоколу інтерфейсів південного напрямку, використовувана версія також є змінним

фактором, який додається у вигляді нових полів і можливостей через останні версії.

Протоколи API північного напрямку використовують для встановлення з'єднання між SDN контролером та існуючими прикладними програмними компонентами. Це з'єднання може встановлюватись з різними цілями, такими як вибірка статистики потоку, керування існуючими та додавання нових потоків, підключення до нових пристроїв чи редагування адміністративних дозволів користувача. Протокол, який використовується для цього з'єднання є відкритим для розробників контролера, так як для цієї мети не визначено конкретного інтерфейсу, і навіть якщо інтерфейс використовує ту ж саму архітектуру (наприклад REST), немає існуючої стандартизації для наданих функцій і відповідної номенклатури API. Існуючі інтерфейси керування контролера безпосередньо пов'язані з інтерфейсами API північного напрямку, так як і ті й інші відповідають за керування та конфігурацію контролера. Також можуть використовуватись інші поширені механізми (наприклад, інтерфейс командного рядку (CLI) або веб-сторінка керування).

Коли говорять про архітектуру контролера, виділяють два її види: розподілену та централізовану. Проблема масштабованості, яка розглядалась в попередньому пункті, має важливе значення в цьому питанні. Хоча централізована архітектура контролера робить задачу реалізації мережі SDN простішою, вона також може стати вузьким місцем мережі, так як не в змозі обробити всі мережеві запити. Таким чином, розподілена архітектура контролера може стати оптимальним рішенням для більш складних мереж, так як робоче навантаження може розподілятися між різними машинами. З іншого боку, підтримка типової глобальної мережі SDN між розподіленими контролерами може створити нові компроміси та додаткові задачі, такі як неузгодженість при підвищенні складності логіки прикладних програмних компонентів.

На сьогоднішній день існує кілька доступних контролерів, які можуть бути встановлені для конфігурації мережі з концепцією SDN.

Одним із перших контролерів, розроблених на базі протоколу OF є *NOX*. Він був написаний на C++ та підтримував версію протоколу OF 1.0. Послідовником даного контролеру є *POX*, написаний на Python. Не дивлячись на те, що ці дві програми активно використовувались користувачами та різними організаціями до 2013 року, сьогодні вони практично не існують, порівняно з іншими рішеннями. Крім того, *NOX* на основі версії протоколу OF 1.3 також був розроблений дослідницьким інститутом CPqD.

Те ж саме трапилось і з OF контролером *Beacon*, розробленим на основі Java з багатопотоковою архітектурою, побудованою на базі сервісної платформи Open Services Gateway Initiative, яка підсилювала його модульність та сервіс-орієнтовані можливості. Незважаючи на надійність цього контролера, він також є застарілим, так як його остання версія була випущена у 2013 році. Подібно до *Beacon*, *Maestro* також є багатопотоковим OF контролером, написаним на Java, розробленим співробітниками університету Rice University. Остання версія контролера була випущена у 2011 році.

Контролер *Ryu* з підтримкою версій протоколу OF 1.0, 1.2, 1.3 та 1.4 пропонує широкий спектр прикладних програмних компонентів, які функціонують на ньому, включаючи концентратор комутації, об'єднувач каналів і т.д. На додачу до його методів інтерфейсу API північного напрямку з архітектурою REST, які використовують для забезпечення керування потоками, він містить базовий набір методів для керування функціональними можливостями QoS екземплярів Open vSwitch.

Контролер на основі мови програмування C під назвою *OpenMul* має архітектуру, яка дозволяє прикладним програмним компонентам працювати та взаємодіяти з контролером через один і той самий простір адрес (через його низькорівневу мову програмування можна взяти цей аспект до уваги),

що сприяє покращенню його продуктивності. Проте, незважаючи на переваги у продуктивності роботи контролера, через те, що він використовує мову C, створення та модифікація програм контролера стає більш складним завданням, у порівнянні з іншими контролерами, які використовують мови програмування більш високого рівня, такі як Java чи Python. OpenMul підтримує версії протоколу OF 1.0, 1.3 та 1.4.

Контролер *Floodlight* на мові Java використовує бібліотеку OF Java, яка підтримує версії протоколу 1.0 та 1.3 з додатковою експериментальною підтримкою версій 1.1, 1.2 і 1.4. Його остання версія була випущена в кінці грудня 2014 року і включала в себе вбудовані програми, такі як пагін інтеграції OpenSteck, віртуальний комутатор, брандмауер та програму Circuit Pusher, яка відповідає за визначення шляхів потоків між хостами.

Opendaylight – це OF контролер, яким керує компанія Linux Foundation за підтримки таких компаній, пов'язаних з комп'ютерними мережами, як Cisco, Brocade, HP, IBM та Intel. Він забезпечує структуровану багаторівневу архітектуру, яка надає декілька вбудованих функціональних можливостей за підтримки відповідних інтерфейсів API північного та південного напрямків. Він підтримує версії протоколу OF 1.0 та 1.3 і, подібно до *Beacon*, є структурованим на базі платформи OSGi, за рахунок чого спрощує керування встановленими і активними функціями. З іншого боку, через таку ж саму високорівневу архітектуру контролера, розробка нових прикладних програмних компонентів для нього є досить складною, так як потребує розуміння усіх необхідних процедур даної платформи. Проте завдяки значній підтримці багатьох організацій, наявності списку розсилки, каналу Internet Relay Chat (IRC), присвяченої контролеру сторінки на Wiki та сайту для запитань і відповідей, досить просто отримати допомогу від інших користувачів та розробників.

В Таблиця 3.1 представлений список існуючих SDN контролерів з версіями протоколу OF, які вони підтримують, використовуваними мовами програмування та назвами компаній-розробників.

Таблиця 3.1

Характеристики існуючих SDN контролерів

Контролер	Платформа	Розробники	Версії OpenFlow
NOX	C++	Nicira	1.0
POX	Python	Nicira	1.0
Beacon	Java	Stanford	1.0
Floodlight	Java	Big Switch Networks	1.0, 1.1, 1.2, 1.3, 1.4
Maestro	Java	Rice University	1.0
Ryu	Python	Ryu SDN Community	1.0, 1.2, 1.3, 1.4
Nox3oflib	C++	CPqD	1.3
OpenDayLight	Java	Linux Foundation	1.0, 1.3
OpenMul	C	Kulcloud	1.0, 1.3, 1.4

3.3.3. Інтеграція SDN та NFV

Інтеграція SDN та віртуалізація мережевих функцій зводить до мінімуму зміни в мережевих елементах, забезпечуючи тим самим плавний перехід на основі потреб оператора. Це дозволяє оновлювати мережеві елементи в деяких частинах мережі, не чіпаючи елементи в інших частинах мережі.

Безпека має першочергове значення для мобільних мереж і її слід розглядати для всіх рівнів, мережевих функцій, а також фізичних і віртуальних елементів. Починаючи від контролера SDN, який має доступ до всієї мережевої архітектури, закінчуючи вузлом, який виконує функції

мережі, система повинна гарантувати критичний рівень безпеки і високої доступності.

Ще однією важливою вимогою є моніторинг мережі. Моніторинг мережі полегшує перевірку і підтвердження Service Level Agreements (SLA), продуктивності (якість обслуговування - QoS), пошуку та усунення несправностей, а також оцінку та оптимізацію використання ресурсів. З одного боку, віртуалізація мережевих функцій встановлює нові вимоги до моніторингу мобільного мережі, але з іншого боку також надає засоби для реалізації передових рішень моніторингу мережі. NFV/SDN дозволяє інтегрувати хмарні інфраструктури, що забезпечують більш високу ступінь свободи щодо розміщення точок вимірювання та гнучкого керування потоками трафіку. Передове і ефективне рішення моніторингу QoS має включати як розподілену (на основі SDN/NFV) систему вимірювання QoS, так і централізовану систему оцінки.

Налаштування сервісів і оптимізація – ще одні вимоги для забезпечення доступності ресурсів. Це може бути зроблено за допомогою оркестратора. Ця вимога може бути розгорнута в SDN мережах, використовуючи прикладні програмні компоненти керування, які мають повне уявлення про конфігурацію мережі. Разом з інформацією про статус від систем моніторингу мережі та збору даних, це дозволяє оркестратору мобільної мережі оптимізувати обслуговування (наприклад, затримку) і/або використовувати менше ресурсів, ніж традиційні мережі. Оркестратор може контролювати кілька мережевих елементів через прикладні програмні компоненти керування. Це дозволяє впроваджувати нові послуги вносячи зміни в оркестратор, тоді як в традиційних мережах для підтримки нових сервісів необхідно оновлювати все обладнання. Мереж на базі SDN повинні співіснувати з наявною архітектурою мережі. Для того, щоб задіяти потенціал SDN, потрібна взаємодія цих двох рішень, наприклад, шляхом введення рівня абстракції і автоматизації в застарілу мережеву частину.

3.4. Віртуалізація в мережах мобільного зв'язку

Віртуалізація відокремлює модель обслуговування системи від її фізичної реалізації з метою підвищення продуктивності обчислень системи (наприклад, обслужити більшу кількість користувачів). Головна перевага віртуалізації полягає в хмарних обчисленнях, в яких технологічно різні системи зібрані в один віртуальний домен (колекція фізичних серверів). Розподіл обчислень на кілька серверів викликає питання про розподіл навантаження на двох рівнях: планування віртуальних машин (VM) і розподілу навантаження. VM планувальник хмарної платформи, в ідеалі, розподіляє обчислення рівномірно між фізичними серверами, одночасно зберігаючи кількість серверів якомога менше, щоб обслуговувати всіх клієнтів. Цей розподіл може бути змінений шляхом міграції віртуальних машин на інші сервери. Єдине обмеження з боку відмовостійкості: віртуальні машини повинні розподілятися з метою пом'якшення наслідків збою обладнання. З іншого боку, балансування навантаження розділяє роботу між аналогічними елементами обслуговування. Якщо VM знаходяться на одній хмарі, навантаження повинно бути рівномірно розподілене на кожній з них. Проте, якщо вони розгорнуті в різних місцях (наприклад, територіально розподілених центрах обробки даних), додаткові вимоги, такі як затримка, змінюють розподіл.

Важливою частиною використання хмарних обчислень є вибір технології, що лежить в основі керування обчислювальної інфраструктурою хмари. Велика кількість обчислень, разом з міркуваннями безпеки, приводить до вибору приватної хмари. Приватна хмара дозволяє розмістити обчислювальні ресурси всередині мережевої архітектури, замість публічної мережі Інтернет, що приводить до зниження затримок і покращує контроль.

Останнім часом досить поширеною є практика використання платформи з відкритим вихідним кодом, з метою уникнення прив'язки до вендору або створення власних функцій, які розширюють можливості платформи.

OpenStack є одним з таких open-source рішень [4]. Рішення підтримує велику кількість гіпервізорів і мережевих технологій. Крім того, OpenStack широко використовується – від невеликих приватних хмар до комерційних рішень, наприклад, Rackspace.

3.4.1. Інтеграція SDN в мобільних мережах

Є кілька робіт, які описують інтеграцію SDN в мобільних мережах [5-9]. Вони пропонують додавати SDN-агенти в елементи мобільної мережі. SoftRAN [5] пропонує централізовану архітектуру в якості альтернативи розподіленому рівню керування, який реалізується в даний час в LTE мережах. Це абстрагує всі базові станції, розгорнуті в географічній зоні в якості великої віртуальної базової станції. Всі фізичні базові станції виглядають як радіоелементи з мінімальною логікою керування. Ці радіо елементи потім керуються централізованим елементом. CellSDN [6] надає детальну класифікацію пакетів комутаторам доступу, яка може бути реалізована в програмному забезпеченні (наприклад, з використанням відкритого VSWITCH). Інші роботи [7-9] визначають, що кожна базова станція має комутатор доступу, який виконує детальну класифікацію пакетів на шляху від мобільного терміналу. Комутатори доступу можуть бути програмними (наприклад, Open vSwitch). Сервер також може мати локального агента, який кешує політики обслуговування для мобільних терміналів, щоб звести до мінімуму взаємодію з центральним контролером.

Інша частина ядра мережі складається з комутаторів, в тому числі кількох шлюзів, підключених до Інтернету. Ці основні комутатори виконують багатовимірну класифікацію пакетів на високій швидкості для кількох тисяч або десятків тисяч правил.

В даний час існують лише кілька поглиблених наукових доповідей, що стосуються мобільних мережевих архітектур, які поєднують в собі поняття хмарних обчислень, SDN та NFV. Перші архітектурні пропозиції - особливо в

контексті хмарних-RAN - включають відображення мережевих функцій, необхідних для інтеграції мобільних мереж з технологією SDN. Це лише функції контролю мобільного зв'язку, тобто MME, HSS, PCRF і рівня керування S/P-GW. Додаткові функції включають в себе транспорт, балансування навантаження, безпеку, політики, білінг, моніторинг, QoE або оптимізацію ресурсів. Ці функції працюють в мобільній Cloud мережі як SDN-прикладні програмні компоненти. При такому підході, призначений для користувача рівень складається тільки із стратегічно розташованих SDN сумісних комутаторів і звичайних комутаторів. SDN комутатори замінюють повністю або частково поточний транспорт мобільної мережі [10-11]. Це консолідована архітектура показана на Рис. 3.3.

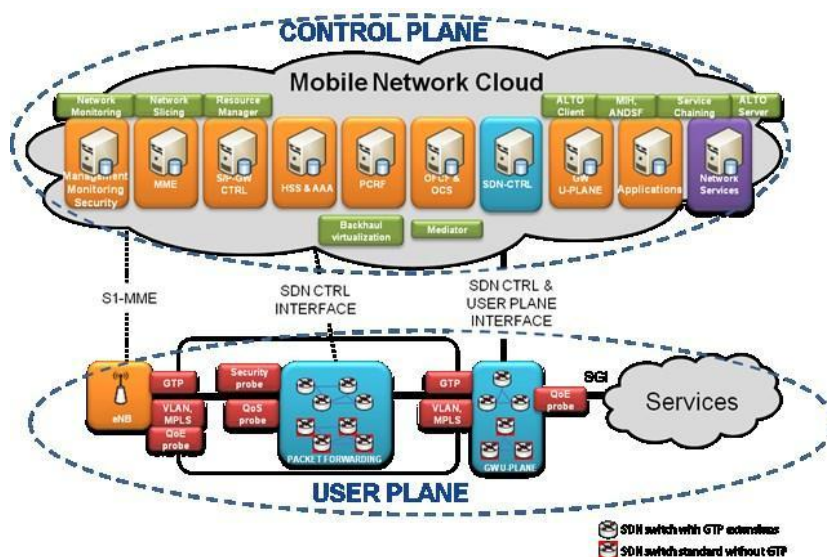


Рис. 3.3 Консолідована архітектура мобільної мережі

Необхідні елементи мережі ЕРС працюють в хмарі. Затримка може вплинути на розташування розгортання деяких обчислювальних вузлів, що працюють віртуально. Деякі стратегічні функції можуть бути розміщені близько до важливих вузлів, створюючи таким чином децентралізовану хмару.

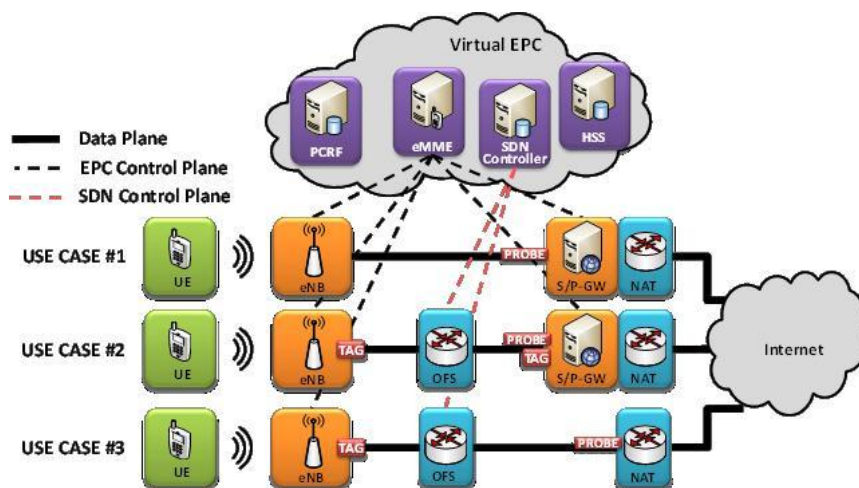


Рис. 3.4 Сценарій міграції з використанням OpenFlow

Рис. 3.4 являє собою 3-кроковий сценарій міграції з використанням OpenFlow як протоколу зв'язку з SDN. Перший випадок використання (тобто UC1) відповідає традиційній LTE архітектурі, зі старими вузлами. Другий випадок використання (тобто UC2) вводить технологію SDN для керування 2-м рівнем маршрутизації на ядрі мобільної мережі, зберігаючи старі вузли. Цей сценарій є гібридним підходом. І, нарешті, третій випадок використання (тобто UC3) зображує повністю сумісну мережу SDN.

3.5. Метод організації ресурсів віртуалізованих мережевих функцій EPC

3.5.1. Підхід до віртуалізації EPC

При віртуалізації архітектури EPC, функції IP Mobile Core можуть бути декомпоновані на віртуалізовані підфункції, які розміщуються на окремих віртуальних машинах. Різні типи віртуальних машин використовують для виконання різних завдань віртуальних підфункцій.

Підхід до віртуалізації EPC архітектури повинен забезпечувати різноманітні аспекти, серед яких можна виділити наступні:

1. Мережеві функції EPC повинні мати хмарно-оптимізовану архітектуру, що дасть можливість найбільш повно реалізувати переваги архітектур NFV/SDN.

2. Масштабованість для забезпечення максимальної операційної гнучкості при розгортанні віртуалізованих функцій EPC у необхідному масштабі.

3. Підтримка потенціалу для послуг мобільного широкосмугового доступу для забезпечення ствлого рівня продуктивності для задоволення кінцевих потреб користувачів.

4. Підтримка кращої доступності і надійності у порівнянні з існуючим пакетного ядра мережі завдяки новим схемам захисту;

5. Підтримка загальних операцій керування мережевими функціями у всій віртуалізованій і фізичній площині EPC.

При цьому серед очікувані переваг від віртуалізації для мобільного ядра можна виділити наступні:

1. Покращена операційна ефективність: NFV інфраструктура буде збільшувати ефективність операційних процедур за рахунок зниження мережових витрат і спрощених операцій.

2. Оптимізована конфігурація та/або топологія мережі за допомогою моніторингу продуктивності: автоматизоване підключення віртуальної машини і оптимізація можуть бути забезпечена використанням політики маршрутизації.

3. Підтримка так званої мульти-оренди: кілька мережових функцій можуть бути сконфігуровані на базі однієї й тієї ж NFV інфраструктури.

4. Скорочення часу виходу на ринок нових послуг: «хмарна» автоматизація дозволяє прискорити впровадження нових сервісів.

5. Введення цільових сервісів на основі географії або місцезнаходження клієнта.

В першу чергу, віртуалізоване ядро EPC, як правило, повинно бути розгорнуто паралельно основному EPC і бути зосередженим на сервісах типу машина-машина та налаштованих сервісах рівня підприємства. Ці види послуг можуть мати певні профілі трафіку, які можуть надати бажані переваги з використанням конкретної конфігурації EPC (і, зокрема,

конфігурації віртуального вузла PGW), і звісно ж таки надати можливість для операторів експериментувати з віртуальним ядром перед впровадженням технології до масового ринку послуг.

Як вже зазначалося вище, розгортання віртуалізованого EPC і пов'язаного з ним надання послуг потребує забезпечення складного оркестрування віртуальних ресурсів як усередині ЦОД, так і у фізичній мобільній транзитній мережі.

З точки зору організації елементів мережі, а зокрема елементів ядра, мережевий сервіс, що забезпечується роботою певного функціонального вузла (або групою вузлів) можна розглядати як граф передачі мережевих функцій (Network Functions, NFs), з'єднаних між собою за допомогою мережевої інфраструктури [18]. Ці мережеві функції можуть бути реалізовані в одній мережі оператора або у взаємодії між різними мережами оператора. Логіка функціонування базової функції мережі обумовлює логіку функціонування сервісу вищого рівня. Отже, функціонування мережевого сервісу в цілому може бути представлене як поєднання логік функціонування блоків, що входять до її складу, які в свою чергу можуть включати окремі мережеві функції, набори мережевих функцій та/або мережеву інфраструктуру. Рис. 3.5 відображає основну схему цього підходу.

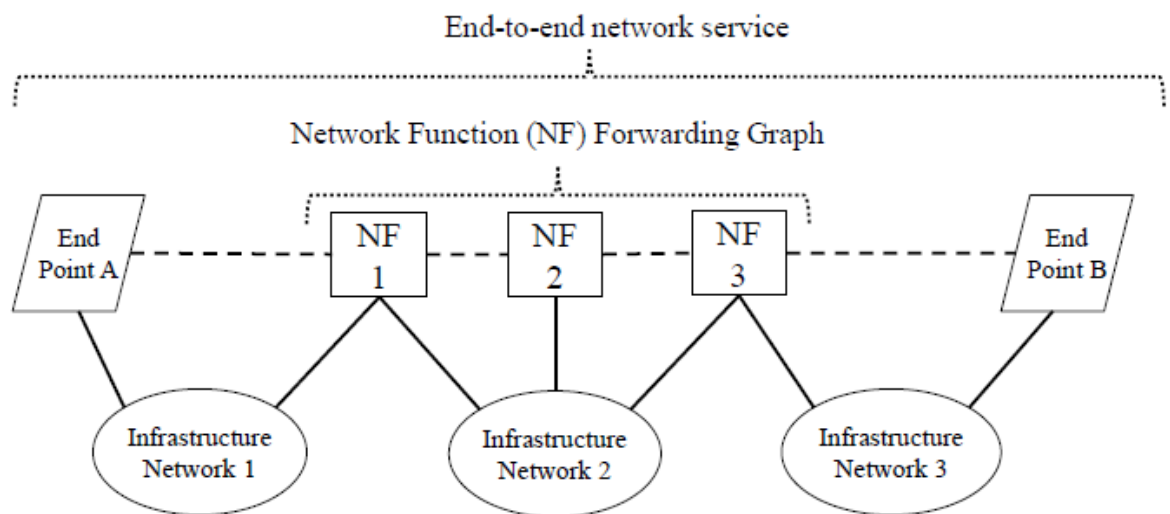


Рис. 3.5 Граф передачі мережевих функцій

Сьогодні, більшість сервісів потребує використання декількох мережових функцій. Оркестрування допомагає інтегрувати і використовувати віртуалізовані мережові функції найбільш зручним і ефективним чином. Система використовує шаблони оркестрування сервісів, які визначають технологічний процес автоматизації для різних функціональних процесів. Окрім цього, архітектура NFV також може забезпечувати формування нового набору вимог до керування операціями в області оркестрування.

Головне питання в полягає у віртуалізації ядра мобільних мереж полягає в тому, якого рівня ефективності вдасться досягти, яка в свою чергу залежить від одного фактора — де будуть розміщені віртуалізовані функції. Впровадження віртуалізації в EPC в цілому повинно бути спрямовано на завдання підтримки необхідного рівня продуктивності та якості обслуговування QoS.

Віртуалізація EPC передбачає, що архітектура NFV буде реалізовувати одну або декілька віртуалізованих мережових функцій (VNFs), а це означає віртуалізацію мережової функції в області базової мережі без віртуалізації.

При цьому VNF може складатися з декількох внутрішніх компонентів і сама по собі автоматично не забезпечувати необхідний функціональний сервіс.

Хмарна інфраструктура передбачає використання віртуальних машин (Virtual Machine) для організації ресурсів віртуалізованих функцій, що дозволить застосувати методи для підвищення доступності ресурсів за допомогою механізмів керування, що автоматично застосовують до екземплярів віртуальних сутностей в хмарній інфраструктурі шляхом використання найбільш ефективного ядра процесора, пам'яті та інтерфейсів, повторної ініціалізації та міграції віртуальних машин. Оскільки кожна віртуальна машина в контексті фізичних ресурсів працює відокремлено, а отже не залежить від інших віртуальних машин і не впливає на їх продуктивність. Ця особливість дає змогу за допомогою обраних інтерфейсів динамічно сконфігурувати ресурси ядра мережі (посилання, топологію

мережі і т.д.), виходячи з факторів необхідної потужності і необхідних транспортних моделей.

Основна ідея, яка пропонується у підході до віртуалізації EPC — це використання декількох менеджерів VNF для забезпечення оперативного контролю в точці розподілу віртуальних функцій в рамках загальної схеми. Зрозуміло, що, в такому випадку, для автоматизації оркестрування, всі операції налаштування, які було передбачено виконувати вручну тепер повинні бути розроблені з урахуванням підтримки машинних форматів опису. Хмарна інфраструктура використовує описи для організації віртуальної мережі, для цього менеджерів VNF здійснюють моніторинг ресурсів в усіх областях і шарах функціональності.

Вище було описано формальний підхід, який пропонується використовувати при віртуалізації ядра мобільної мережі. Для ефективного впровадження віртуалізації необхідно визначити, які функції доцільно віртуалізовувати при запропонованому підході. Для цього розглянемо схему організації функціональних вузлів ядра мережі LTE (Рис. 3.6) та призначення елементів [19].

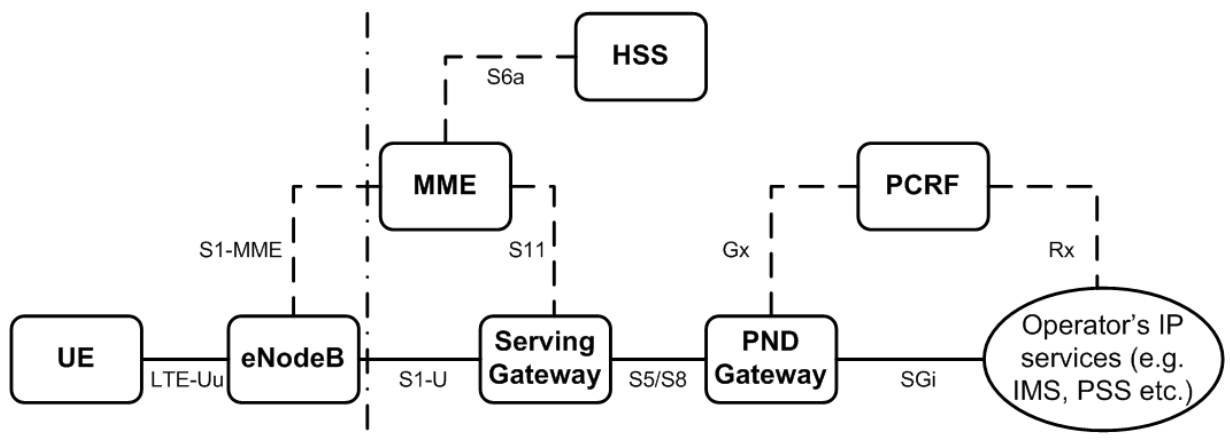


Рис. 3.6 Схема функціональних вузлів EPC

Базова станція (EnodeB) в мережі LTE виконує наступні функції: керування радіоресурсами (Radio Resource Management, RRM), вибір блоку керування мобільністю (MME) при включенні в мережу користувацького терміналу, маршрутизація в площині користувача пакетів даних у напрямку

до обслуговуючого шлюзу (SGW), диспетчеризація і передача інформації виклику і мовної інформації, отриманої від блоку керування мобільністю, диспетчеризація і передача повідомлень PWS (Public Warning System, система тривожного сповіщення), отриманих від MME.

MME — це основний елемент керування в мережі LTE. Він здійснює лише функції керування і не працює з даними користувачів. Має безпосередній зв'язок з терміналом користувача (User Entity, UE) через протокол сигналізації поза рівнем доступу (Non Access Stratum, NAS).

Функції, що виконуються MME, можна розділити на наступні два набори: керування потоком (на пред'явника керування) та керування підключеннями (керування з'єднання).

Загалом же до функціоналу MME відноситься сигналізація між мережею EPC і UE, сигналізація хендоверу між різними мережами, вибір PGW і SGW, аутентифікація при реєстрації UE в мережі, керування каналами на інтерфейсах до інших елементів мережі.

SGW призначений для обробки і маршрутизації пакетних даних, що надходять з/в підсистему базових станцій.

SGW маршрутизує і направляє пакети з даними користувачів, в той же час виконуючи роль вузла керування мобільністю (mobility anchor) для користувацьких даних при хендовері між базовими станціями, а також вузла керування мобільністю між мережею LTE і мережами з іншими технологіями 3GPP. Коли UE вільний і не зайнятий викликом, SGW підключає спадний канал даних (DownLink, DL) і виконує пейджинг, якщо потрібно передати дані по DL в напрямку UE. Він керує і зберігає стан UE (наприклад вимоги по пропускній здатності для IP-сервісів, внутрішню інформацію щодо мережевої маршрутизації).

PGW забезпечує з'єднання від UE до зовнішніх пакетних мереж даних, будучи точкою входу і виходу трафіку для UE. UE може мати одночасно з'єднання з більш ніж одним PGW для підключення до декількох мереж.

PGW виконує функції захисту, фільтрації пакетів для кожного користувача, підтримку білінгу, узаконеного перехоплення і сортування пакетів.

Виходя з типу функціонал вузлів сервісного і пакетного шлюзів може бути розділений на дві частини: площину керування і площину даних. Таке розділення дозволяє перемістити функції площини керування (SGW-Control та PGW-Control) до централізованої хмарної платформи. При цьому VNF менеджери будуть обслуговувати віртуалізовані функції SGW та PGW, що працюють на відокремлених віртуальних машинах, і відповідати за оркестрацію життєвого циклу VNF. Рис. 3.7 схематично ілюструє підхід до організації EPC, що пропонується.

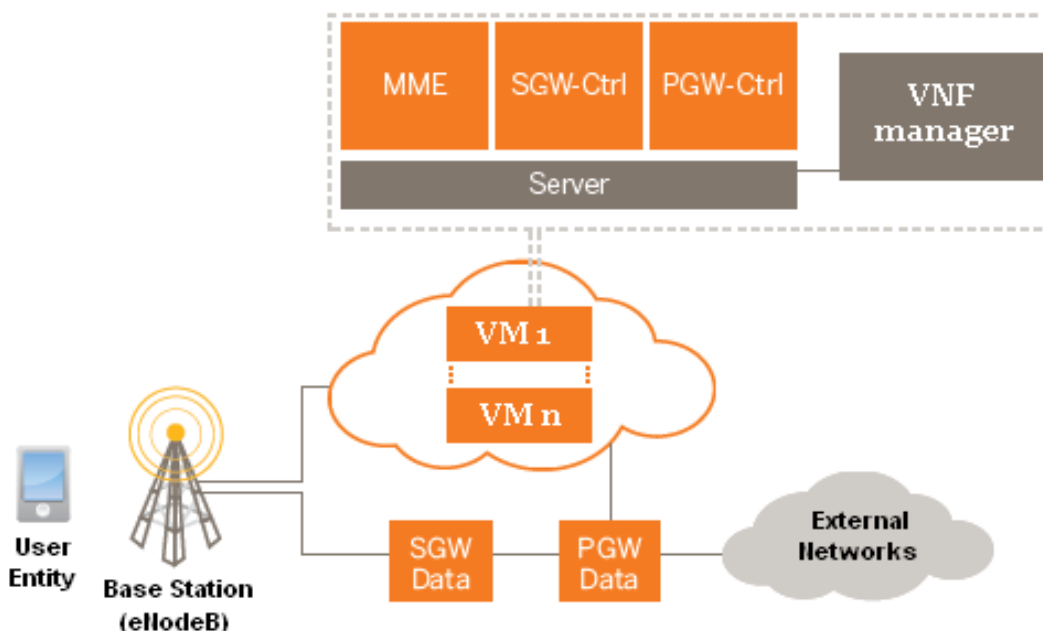


Рис. 3.7 Підхід до організації EPC з виділенням функціоналу SGW-Ctrl та PGW-Ctrl

Підхід до організації EPC, який пропонується, передбачає, що робота мережі в цілому буде ґрунтуватися на декількох серверах або центрах обробки даних. При цьому, пропонується організувати SGW/PGW і MME функції на спільно на відособлених комплексах центрів обробки даних.

Взаємодія вузлів мережеских функцій забезпечується інкапсульовано стандартними інтерфейсами (s6s для взаємодії MME-HSS, s11 для MME-S/P-GW і s5 для окремих S/P-GW вузлів).

Таким чином, підхід дозволяє масштабувати S/P-GW і мережеві функції MME відповідно до їх власних потреб в ресурсах. Це ефективно, коли, наприклад, може виникнути необхідність збільшити обсяг ресурсів у площині користувача, не впливаючи на рівень керування, і навпаки.

Також зазначимо, що при динамічній реконфігурації віртуальної функції мережі унаслідок збою або перевантаження з автоматичним або ручним режимом керування, сеанси та/або сесії керування повинні бути оброблені відповідним чином для досягнення необхідної надійності послуг.

3.5.2. Організація функціональних вузлів мережі LTE

PCRF — функціональний елемент в мережах зв'язку рекомендований групою 3GPP, основною функцією якого є впровадження політик обслуговування абонентів, наприклад, дозвіл/заборона сервісів або встановлення параметрів якості обслуговування, тобто застосування певних правил до корисного інформаційного потоку. Крім того, функціональний елемент PCRF встановлює правила тарифікації залежно від різних умов, таких як: параметри абонентського профілю, час доби, місця розташування абонента, обсяг спожитого трафіку та інших.

На сьогоднішній день функції PCRF реалізовані на програмно-апаратному забезпеченні, що обмежує можливості його масштабування. При цьому всі інформаційні потоки виявляються прив'язаними до топології мережі оператора, так як для забезпечення підрахунку кількості трафіку і для контролю якості обслуговування всі інформаційні потоки повинні взаємодіяти з блоком PCRF. Саме тому після появи концепції програмно-керованих мереж і віртуалізації мережевих функцій, стало можливим застосування даних концепцій для удосконалення роботи функціонального блоку PCRF в рамках архітектури EPC.

Запропонований метод має на меті віртуалізацію деяких функцій PCRF, яка полягає в тому, що ряд функцій реалізується на базі віртуальних машин

розміщених в хмарі. При цьому стає можливим масштабування самого PCRF при розгортанні мережі без необхідності організації окремих реальних фізичних серверних потужностей.

Організація взаємодії віртуальних машин PCRF з низкою вузлів мережі зв'язку є завданням, яке заслуговує особливої уваги. При цьому важливим питанням є розподіл функцій між віртуальною PCRF та іншими функціональними елементами системи 3GPP які пропускають через себе інформаційний потік.

Функціональні вузли мережі, з якими взаємодіє PCRF, обслуговують певний сегмент мережі, відповідно пропонується організовувати віртуалізацію вузлів, що відповідають за один сегмент на одному хмарному «сервері» для оптимізації взаємодії. Для оптимізації адміністрування функціональні вузли-контролери, що обслуговують різні сегменти мережі будуть організовані на окремих віртуальних машинах. Саме вузли-контролери функціональних елементів PCRF будуть здійснювати взаємодію з іншими елементами архітектури сервера оператора.

Розглянемо детально функціональні елементи і інтерфейси, з якими взаємодіє PCRF [20].

Перший елемент, Policy and Charging Enforcement Function (PCEF) — функціональний елемент в 3GPP мережах зв'язку, який здійснює застосування Policy and Charging Control (PCC)-правил, отриманих від PCRF, до трафіку, що проходить через нього. Здійснює тарифікацію цього трафіку в системі тарифікації оператора зв'язку OCS/OFCS.

Даний компонент в якості функціонального блоку PGW виноситься в хмарну інфраструктуру. Взаємодія хмарного компонента PCEF у складі PGW-Control з PCRF здійснюватиметься за Gx-інтерфейсом, який використовує протокол Diameter і призначений для надання службових даних по реалізації Flow Based Charging — FBC правил білінгових розрахунків з абонентами. По цьому інтерфейсу PCEF передає на PCRF інформацію, необхідну для прийняття PCC-рішень: ідентифікатор абонента, інформацію

щодо місцезнаходження і часового пояса, в якому знаходиться абонент, IP-адресу пристрою, з якої здійснюється доступ, параметри каналу, та інші.

На реальному фізичному обладнанні буде працювати варіант «тонкого клієнта» PGW — PGW-Data, завдання якого полягає у фільтрації трафіку і застосуванні правил в залежності від інструкцій, одержуваних від PGW-Control. Взаємодія клієнта і хмарної реалізації PGW здійснюється за допомогою інтерфейсу s5.

Функції фільтрації пакетів по користувачам і законного перехоплення трафіку (позначимо даний набір функцій як F1) здійснюються на боці «тонкого клієнта» PGW, в хмарну інфраструктуру виноситися функціонал розподілу пулу IP-адрес для пристроїв користувачів UE та функціонал PCEF (F2).

Призначення PCC-правил, якими обмінюються між собою елементи мережі PCRF і PCEF — поділ фізичного потоку даних (IP-CAN) на логічні сесії SDF (Service Data Flow), визначення того до яких прикладних програмних компонентів і послуг відноситься трафік, надання параметрів QoS та інформації для тарифікації. Використовується два типи PCC-правил: динамічні PCC-правила, які передаються з PCRF на PCEF через Gx-інтерфейс і визначені на PCEF. Визначені правила можуть бути активізовані або PCRF, або самим PCEF.

Наступний елемент, який буде віртуалізовано — Bearer Binding and Event Reporting Function (BBERF). Це функціональний елемент в 3GPP мережах зв'язку, який здійснює нотифікацію PCRF про встановлення сесії з посилкою ідентифікатора абонента і додаткових параметрів для коректного визначення QoS-правил обслуговування. Функціонал даного компонента буде винесено в хмарну інфраструктуру в якості функціоналу вузла SGW-Control.

«Тонкий клієнт» SGW-Data буде виконувати наступні функції: базову маршрутизацію пакетного трафіку і перехоплення пакетного трафіку (F3), а також функціонал "якірної" точки (точки об'єднання трафіку) для хендвера

між базовими станціями NodeB всередині однієї мережі доступу в зоні обслуговування базових станцій згідно набору правил та інструкцій (F4), що надходитимуть від хмарної «серверної» частини SGW-Control.

Елемент SGW-Control в свою чергу забезпечує виконання таких функцій: "якірна" точка для хендовера між різними мережами доступу стандартів LTE/LTE і LTE/UMTS (F4), і обробка функціоналу BBERF (F5).

Взаємодія з PCRF здійснюється вузлом SGW-Control (також як і у випадку з PGW-Control), але по інтерфейсу Gxx. Взаємодія між компонентами SGW-Control і SGW-Data здійснюється по інтерфейсу s5.

Існує, також, ряд вузлів, функціонал яких пропонується розміщувати в хмарній інфраструктурі (з поділом сервером за сегментами мережі). Серед них можна виділити:

1. Traffic Detection Function (TDF, F6) — функціональний елемент в 3GPP мережах зв'язку, який здійснює визначення трафіку певних прикладних програмних компонентів і нотифікацію про нього PCRF. Залежно від отриманих правил здійснює пропуск даного трафіку абоненту, перенаправлення і обмеження швидкості. Взаємодіє з PCRF по інтерфейсу Sd, який використовується для встановлення Application Detection and Control (ADC) правил керування параметрами трафіку конкретних програм.

2. User Data Repository (UDR, F7) — функціональний елемент, який здійснює зберігання даних користувачів. Взаємодіє з PCRF по інтерфейсу Ud, який використовується для отримання/зміни профілів, в яких зберігається інформація про сервіси, що доступні абоненту, параметрах QoS та інших, необхідних для прийняття PCC-рішень. Також інтерфейс Ud використовується для організації передплати та отримання нотифікацій про зміни у профілях абонентів.

3. Application Function (AF, F8) — функціональний елемент, який надає опис потоку даних сервіса і здійснює інформування про необхідні сервісу ресурси. Взаємодіє з PCRF по інтерфейсу Rx.

4. OCS (F9) — сервер кредитного контролю в режимі реального часу, який здійснює тарифікацію послуг, контролює баланс абонента, обробляє інформацію про нарахування та списання коштів на балансі абонента, застосовує знижки, здійснює підрахунок обсягу спожитих послуг. Взаємодіє з PCRF по інтерфейсу Sy, який використовується для обліку обсягів спожитих послуг і нотифікації про подолання порогів лічильників з OCS на PCRF. Крім PCRF, OCS взаємодіє з PCEF по інтерфейсу Gy за допомогою якого здійснюється тарифікація послуг.

Функціонал PCRF, який буде винесено в хмарну інфраструктуру позначимо як F10.

На Рис. 3.8 представлена запропонована схема організації функціональних вузлів мережі (функції, що відмічені червоним кольором віртуалізовані в «хмарну» інфраструктуру).

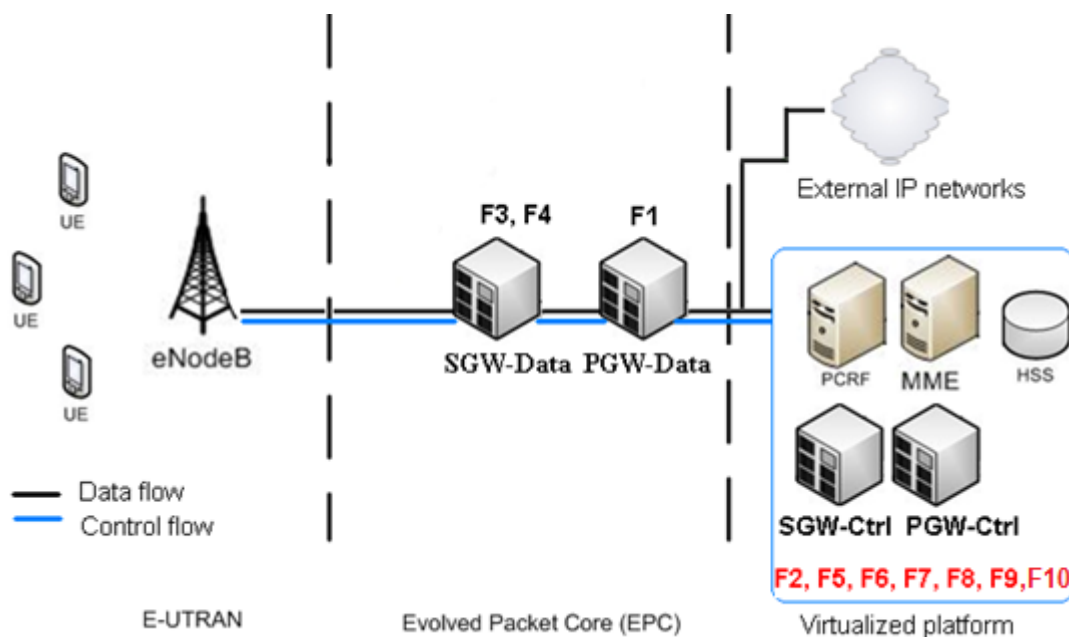


Рис. 3.8 Схема організації функціональних вузлів мережі

3.5.3. Передача повідомлень управління при віртуалізації EPC

У ході взаємодії компонентів мережі LTE мова йде про логічні канали, які надають послуги середнього рівня управління доступом в межах структури протоколу LTE. Логічні канали по типу інформації, що

передається, діляться на логічні канали управління та логічні канали трафіку. Логічні канали управління використовуються для передачі різних сигнальних та інформаційних повідомлень. По логічним каналам трафіку передаються дані користувачів.

Підхід до віртуалізації EPC, що розглядається, дає можливість: по-перше, спростити організацію логічних каналів, так як при даному варіанті організації, по суті, існують тільки логічний канал трафіку UE-Base station-SGW-Data-PGW-Data з виходом на пакетні мережі для обміну даними між різними ділянками мережі і аналогічний логічний канал управління, який використовує вихід на пакетні мережі для організації взаємодії з хмарною віртуалізованою архітектурою; по-друге, максимально уніфікувати обладнання, що здійснює фільтрацію трафіку (SGW і PGW) за допомогою обмеження його функціоналу; і, по-третє, надає можливість резервування серверів віртуалізованих функцій.

Рис. 3.9 ілюструє схему встановлення підключення пристрою користувача до мережі LTE при реалізації методу організації ресурсів EPC.

В цілому послідовність взаємодії в мобільній мережі передачі даних при запропонованому варіанті віртуалізації можна описати наступним чином:

1. Абонент починає сесію передачі даних (Запит на підключення RRC). Дані від базової станції надходять до віртуалізованого блоку MME (S1 Setup) і базової станція розпочинає приєднання UE (UE Attach).
2. MME надсилає до HSS запит на авторизацію даних (Authorization Data Request).
3. У хмарній віртуалізованій інфраструктурі функціонал BBERF надсилає до PCEF запит на створення сесії для пропуску трафіку (IP-CAN) (блоки розміщені на одній віртуалізованій платформі).
4. PCEF формує запит по інтерфейсу Gx і посилає його на PCRF. Запит полягає у формуванні Diameter CCR (Credit-Control-Request) запиту з інформацією про абонента і запитуваних послугах.

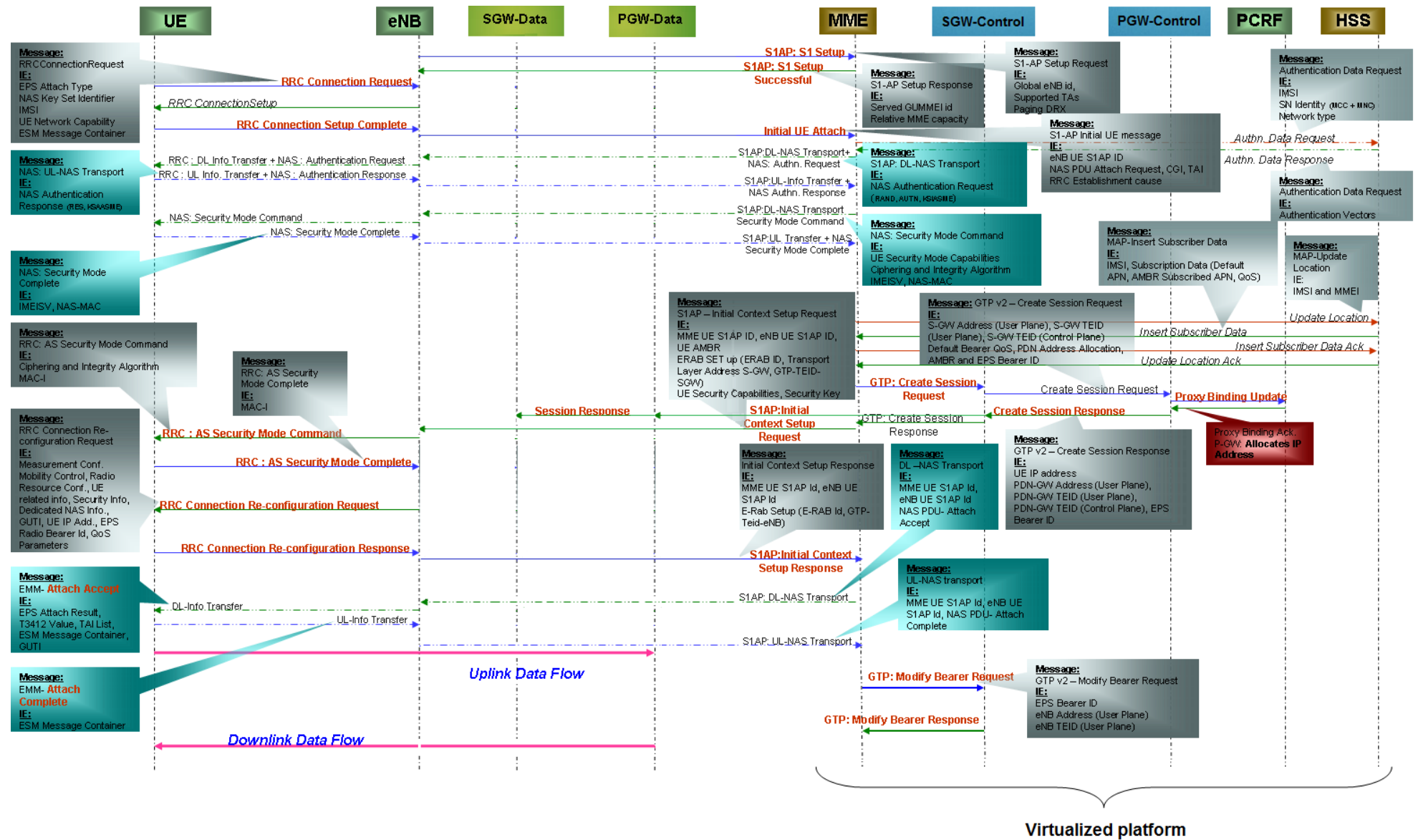


Рис. 3.9 Приклад обміну повідомленнями керування при віртуалізації ядра мережі

5. PCRF здійснює запит профілю абонента по інтерфейсу Ud.
6. Отримує профіль з параметрами послуг абонента.
7. Здійснює підписку на нотифікацію про зміни профілю.
8. PCRF приймає PCC-рішення про можливість надання послуг абоненту і з якими параметрами якості. Формує PCC-правила, які відправляє на PCEF по інтерфейсу Gx. Це полягає у формуванні Diameter CCA (Credit-Control-Answer) відповіді з включеним набором PCC-правил.
9. При отриманні відповіді PCEF встановлює сесію кредитного контролю з OCS по інтерфейсу Gy за допомогою обміну повідомленнями Diameter CCR/CCA.
10. PCEF дозволяє встановлення IP-CAN сесії і передає відповідні повідомлення на SGW-Control і PGW-Control.
11. SGW-контроль надсилає повідомлення Initial Context Setup на PGW-Data і Session Response (відповідь сесії) на SGW-Data.
12. Потік трафіку (Service Data Flow) починає проходити між пристроєм абонента і зовнішніми мережами зв'язку.
13. Через деякий час абонент завершує сесію передачі даних і BBERF посилає на PCEF запит на розрив IP-CAN сесії.
14. PCEF здійснює завершення Diameter сесій на PCRF по інтерфейсу Gx. Завершення сесій полягає також в обміні повідомленнями Diameter CCR/CCA.
15. PCEF здійснює завершення Diameter сесій на OCS по інтерфейсу Gy.

Висновки

1. Розглянуто основні вимоги підходу до віртуалізації EPC та функціональні особливості вузлів ядра мобільної мережі, які впливають на архітектуру віртуалізованої платформи. Прийнято рішення щодо використання декількох менеджерів VNF для забезпечення оперативного контролю в точці розподілу віртуальних функцій.

2. Запропоновано метод організації ресурсів ядра мобільних мереж EPC, який полягає у віртуалізації значної частини функцій, які відповідають за забезпечення виконання білінгових операцій, та уніфікації обладнання, що здійснює фільтрацію трафіку (вузли SGW-control і PGW-Control), що дозволяє підвищити гнучкість керування мережею мобільного зв'язку.

3. Розроблена архітектура локальної мережі оператора мобільного зв'язку дозволяє оцінити переваги та вузькі місця віртуалізації мережевих функцій, та визначити особливості забезпечення показників якості обслуговування у гетерогенному телекомунікаційному середовищі.

4. Вдосконалено принцип керування телекомунікаційною системою на відміну від принципу програмно-керованих телекомунікаційних мереж; функції керування процесом обслуговування абонента, а саме: пошук абонента, пошук фізичних елементів, які беруть участь у процесі передачі, і передача керівного впливу на відповідні фізичні елементи, переносяться у хмару, всі підсистеми мобільного зв'язку є керованими із контролерів розташованих у дата центрі, взаємодія між контролерами підсистем з метою керування відбувається лише в середині дата центру, що дозволить зменшити кількість службових потоків по телекомунікаційній мережі.

РОЗДІЛ 4

РЕЗЕРВУВАННЯ РЕСУРСІВ У ВІРТУАЛІЗОВАНОМУ СЕРЕДОВИЩІ ДЛЯ ОБСЛУГОВУВАННЯ НАДМІРНОГО ПОТОКУ НАВАНТАЖЕННЯ НА ЕЛЕМЕНТИ ГТС

Основна задача даного принципу полягає в тому, щоб обмежити вхідний потік на основний вузол обслуговування і розрахувати параметри додаткового вузла обслуговування.

Для вузла обслуговування визначається максимальну інтенсивність вхідного потоку, який може бути обслужений із заданими параметрами ефективності, для решти потоку визначається оптимальна інтенсивність обслуговування та відповідні ресурси альтернативного вузла обслуговування, які забезпечують ефективне обслуговування службових потоків (Рис. 4.1).

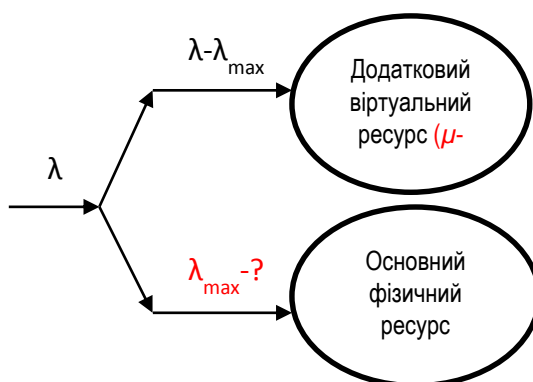


Рис. 4.1 Схема резервування ресурсів у віртуалізованому середовищі

4.1. Процедура гарантування заданої якості обслуговування

Принцип динамічного контролю якості полягає у наступному: значення затримки у обслуговуванні заявки на встановлення з'єднання (роз'єднання, відновлення) порівнюється із політикою якості обслуговування відповідного абонента. Якщо показник не відповідає, тоді послідовно порівнюються показники якості у віртуальних вузлах та каналах зв'язку віртуальної мережі із пороговими значеннями відповідних політик збережених у підсистемі PCRF. Даний принцип аналізує такі кількісні показники ефективної роботи системи, як: час затримки заявки службового потоку у віртуальному вузлі та

ймовірність втрати запитів у вузлі обслуговування. Вузол обслуговування - це віртуальна машина яка виконує функції керування мережевого вузла.

Після того, як було виявлено де саме є проблема зниження показників ефективності обслуговування Приймаються міри: Якщо проблема у часі передачі між вузлами обслуговування, то рекомендується зробити реконфігурацію системи, а саме змінити розміщення віртуальних вузлів у фізичних вузлах гетерогенної структури датацентрів. Якщо проблема ідентифікована у одному вузлі обслуговування, тоді рекомендовано збільшити кількість ресурсів обслуговування. Якщо спостерігається зниження показників якості обслуговування у групі вузлів зв'язаних інтерфейсів, наприклад які утворюють єдине ядро мережі ЕРС, тоді рекомендовано обмежити потік заявок які направляються на обслуговування відповідного ядра. Для цього рекомендовано розрахувати інтенсивність навантаження на групу вузлів. Алгоритм процедури наведений на Рис. 4.2.

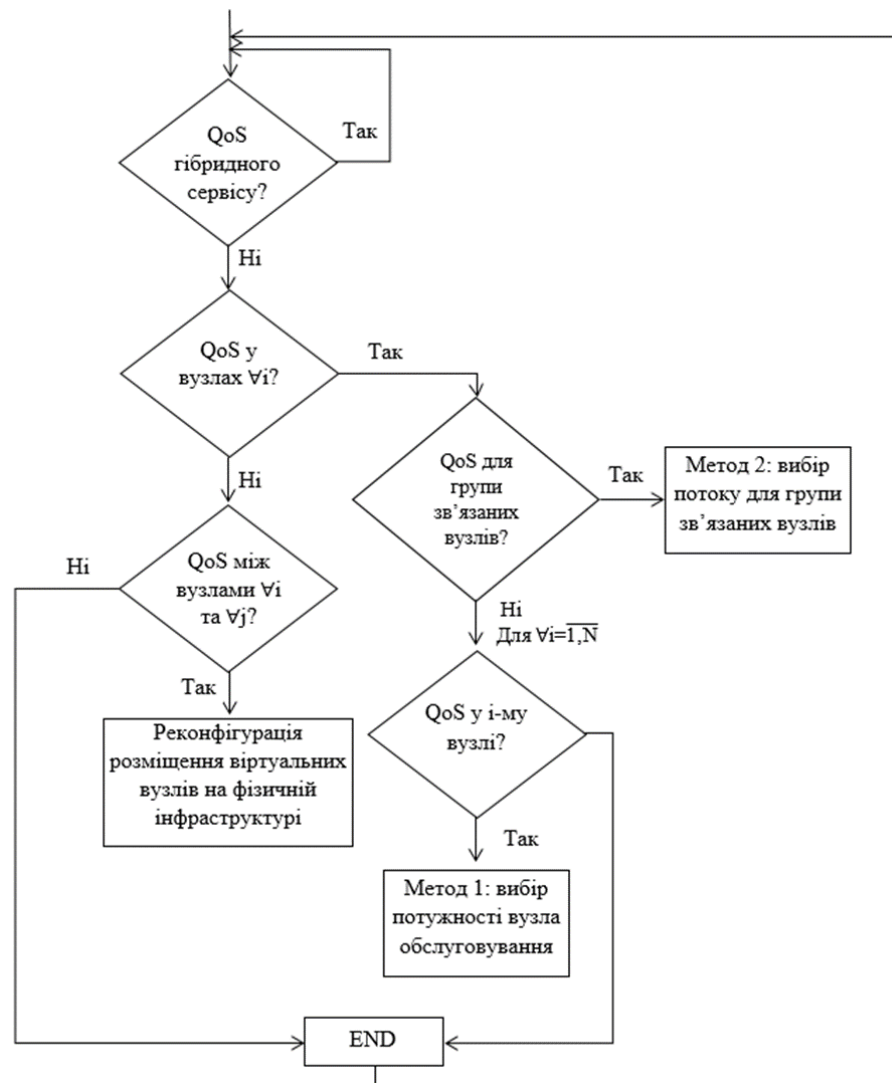


Рис. 4.2 Процедура гарантування заданої якості обслуговування

Для реалізації принципу динамічного контролю якості потрібна модифікація підсистем системи PCRF. Підсистема «Єдине сховище політик» розширюється, додаються наступні політики відносно показників якості обслуговування потоків керування:

1. Допустимий час затримки заявки службового потоку у віртуальному вузлі.
2. Допустимі втрати запитів у віртуальному вузлі
3. Допустимий час обслуговування запитів у групах віртуальних вузлів які забезпечують заданий сервіс.
4. Допустимі затримки при передачі між вузлами обслуговування
5. Значення допустимих затримок доставки керівного впливу на мережеві вузли.

Розширена підсистема зображена на Рис. 4.3.

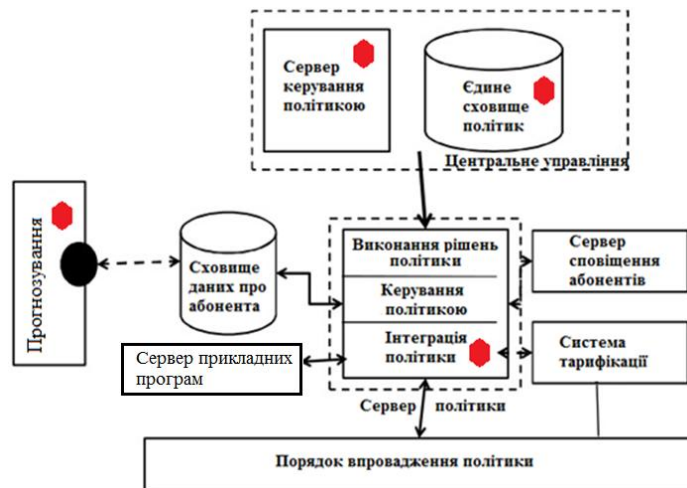


Рис. 4.3 Модифікація підсистеми PCRF

- Підсистема «Керування політикою» формує групу вимог до виконання набору політик по відношенню до різних потоків керування.
- Підсистема «Сервер політики» виявляє проблему невідповідності поточних показників якості заявленим політикам обслуговування відповідного абонента.
- У підсистемі «Сервер прикладних програм» реалізовані програмні модулі в яких виконуються обчислення відповідно до запропонованих методів. Вихідними даними для методів є статистичні дані отримані від системи моніторингу та дані про політики, які забезпечуються для відповідних абонентів.
- Підсистема «Сховище даних про абонента» доповнюється інформацією про віртуальні вузли, або створюється окрема база даних

інформації про статистику функціонування віртуальної мережі обслуговування, де збирається інформація про потоки заявок на обслуговування, статистика відносної залежності інтенсивності обслуговування від ресурсів обслуговування диференційовано для кожного типу запитів.

Принцип динамічного контролю якості обслуговування потребує нових процедур: необхідно організувати взаємодію системи керування мобільного зв'язку із системою керування віртуалізованими ресурсами (Рис. 4.4).

Контроль якості виконання керівних процедур оцінюється на рівні абонентського пристрою:

Абонентський пристрій фіксує час затримки у виконанні службових процедур, а саме час від моменту ініціалізації з'єднання до моменту початку передачі даних, та передає підсистемі PCRf.

PCRf отримує цю інформацію від абонента та аналізує на сервері політик; у підсистемі виконання рішень політики порівнює отримані дані на відповідність обраної політики, яка відповідає абоненту та зберігається у «Сховищі даних про абонента».

Якщо значення параметрів затримок не відповідають політиці, PCRf робить запит до підсистеми «Оркестратор» для визначення групи вузлів i , які обслуговують відповідного абонента.

Підсистема «Оркестратор» надсилає номери вузлів, які обслуговують абонента, розташованого у заданій місцевості. PCRf відправляє підсистемі «Моніторинг хмарних прикладних програм» запит для отримання інформації про час затримки та показники втрат у вузлах i , а також інформацію про затримки між вузлами обслуговування. Система моніторингу хмарних прикладних програм збирає інформацію відносно показників затримки та втрат гібридних сервісів, які обслуговуються у вузлах віртуальної мережі. Дані відносно групи вузлів обслуговування i передають у PCRf, де реалізовано принцип динамічного контролю якості обслуговування гібридних сервісів. Відповідно до керівних рішень, підсистема PCRf направляє запити:

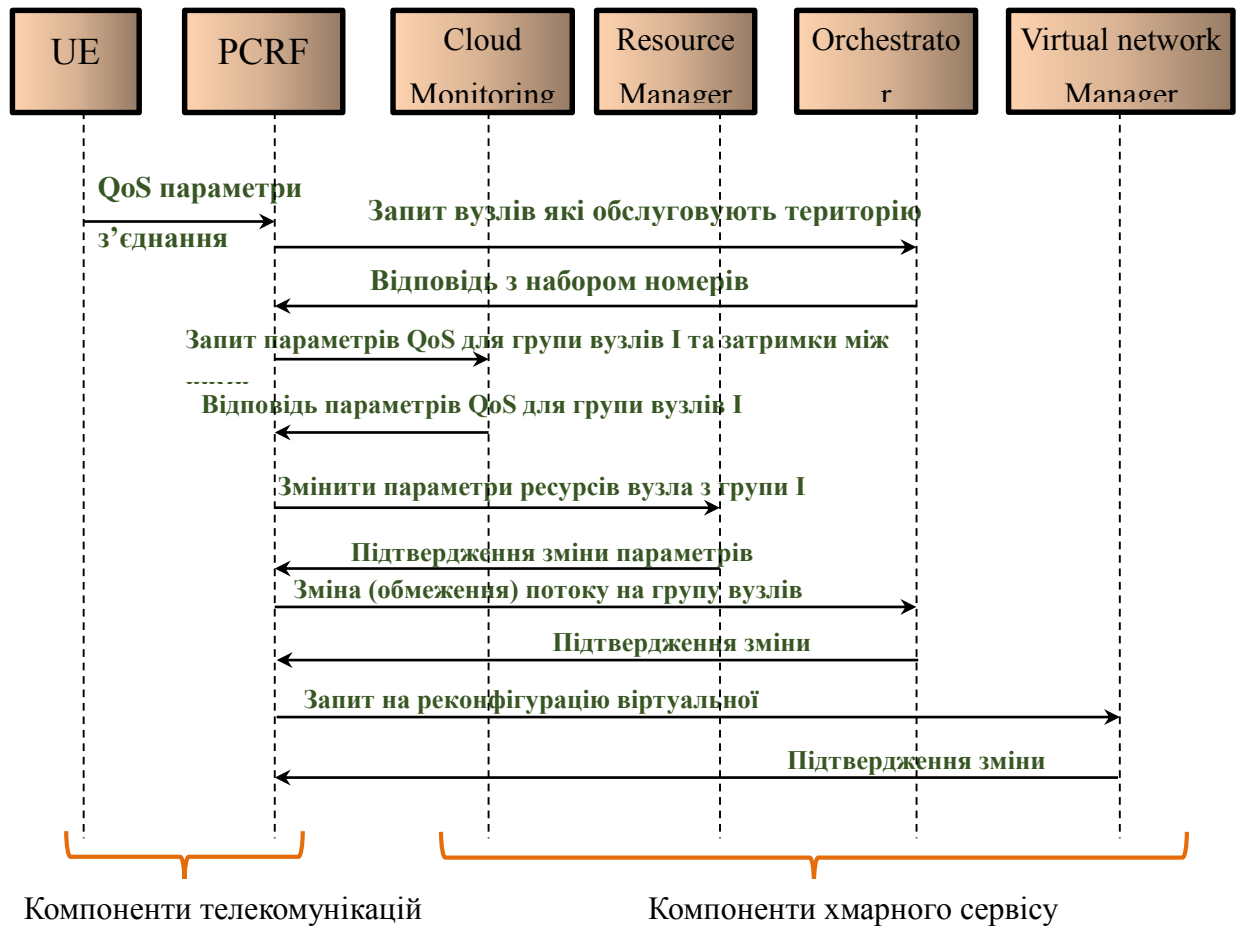


Рис. 4.4 Взаємодія системи керування мобільного зв'язку та системи керування віртуалізованим середовищем

- на реконфігурацію віртуальної мережі, до «Менеджера віртуальної мережі»;
- на реконфігурацію ресурсів до «Менеджера ресурсів»;
- на зміну потоків обслуговування до «Оркестратора» потоків по віртуальній мережі.

При реалізації принципу динамічного контролю якості обслуговування задіяно більшість підсистем системи PCRF.

4.2. Задача формування вхідного потоку навантаження

Задача формування вхідного потоку навантаження також має широке застосування у сучасних телекомунікаційних системах, які обслуговують великі об'єми нестаціонарних потоків запитів від абонентів на надання гібридних сервісів. Ця задача буде корисною не тільки при коректуванні

навантаження на групу віртуальних вузлів зв'язаних інтерфейсів, але й при плануванні потоків на підсистеми обслуговування.

Для розрахунку потоку, який спрямовується на вузол обслуговування, необхідні дані про кількість заявок, які одночасно обслуговуються у вузлі обслуговування, інтенсивність обслуговування заявок у системі, кількість узагальненого обчислювального ресурсу, який використовується для обслуговування однієї заявки, а також загальну кількість ресурсу у вузлі і вимоги до показників якості обслуговування запитів.

Вхідні дані

- n – кількість заявок, які можуть одночасно обслуговуватися у вузлі обслуговування.

- μ – інтенсивність обслуговування заявки,

- ν_k – кількість узагальненого обчислювального ресурсу, який використовується для обслуговування однієї заявки при одночасному обслуговуванні k заявок $k = \overline{1, n}$.

- V – доступний об'єм узагальненого ресурсу, який спільно використовується заявками.

- m – допустима кількість запитів у черзі на обслуговування.

- R – відсоток заявок, які обслуговуються у системі не більше допустимого часу затримки, визначається експертами

- l – кількість запитів у черзі, до досягнення якої блокується надходження запитів до системи відповідно до алгоритмів раннього упередження перевантажень

Вихідні дані

- λ - рекомендоване значення для інтенсивності вхідного потоку, що буде направлено на обслуговування у n -канальний обслуговуючий пристрій.

Розроблена математична модель задачі формування вхідного потоку розрахована на дисципліну обслуговування запитів за алгоритмом раннього

виявлення перевантажень, враховує ергодичний розподіл ймовірностей одночасного перебування у системі k -заявок:

$$p_0 = \frac{1-\beta^{m-l}}{A_n(\alpha, \beta)}, \beta \neq 1, \alpha = (\lambda/\mu), \beta = (\lambda/n)$$

$$p_k = \frac{\alpha^k}{k!} p_0 \quad (k = \overline{1, n})$$

$$p_{n+k} = \frac{\alpha^n}{n!} \beta^k p_0 \quad (k = \overline{1, l})$$

$$p_{n+k} = \frac{\alpha^n}{n!} \frac{\beta^k - \beta^m}{1 - \beta^{m-l}} p_0 \quad (k = \overline{l+1, m-1})$$

$$p_{n+m} = \frac{\alpha^n}{n!} \frac{(1-\beta)\beta^m}{A_n(\alpha, \beta)} \quad (k = \overline{n+l+1, n+m-1})$$

$$A_n(\alpha, \beta) = (1 - \beta^{m-l}) \sum_{k=0}^n \frac{\alpha^k}{k!} + \frac{\alpha^n}{n!} \left(\frac{\beta - \beta^{m-l+1}}{1 - \beta} - (m-l)\beta^{m+1} \right)$$

Була запропонована математична модель пошуку максимально допустимого навантаження на систему для якого вистачить ресурсів обслуговування і з заданою ймовірністю час обслуговування не буде перевищувати допустимого значення.

$$\lambda \rightarrow \max$$

$$\begin{cases} 4 * \left(\sum_{i=1}^n i v_k^g p_i + n v_k^g \sum_{i=n+1}^s p_i \right) \leq V^g, g = \overline{1, G} \\ \sum_{i=1}^s p_i \leq R \end{cases}$$

де s – номер стану, розрахований за ЦГТ на основі вимог до якості обслуговування.

Задача вибору потужності вузла обслуговування є задачею нелінійного програмування. В процесі моделювання використовувалася функція Matlab `fmincon`, яка зупиняється при наближенні до оптимального значення із заданою похибкою.

4.3. Задача вибору потужності вузла обслуговування

Задача вибору потужності вузла обслуговування набуває особливої ваги при плануванні процесів обслуговування гібридних телекомунікаційних сервісів, оскільки вирішує проблему розрахунку кількості обчислювальних ресурсів для обслуговування нестационарного потоку.

Визначення NFV-ресурсів було наведено у специфікації ETSI GS NFV 003. До ресурсів обслуговування відносяться оперативна та дискова пам'ять, процесорний час із розрахунку однопоточного процесору (Рис. 4.5).

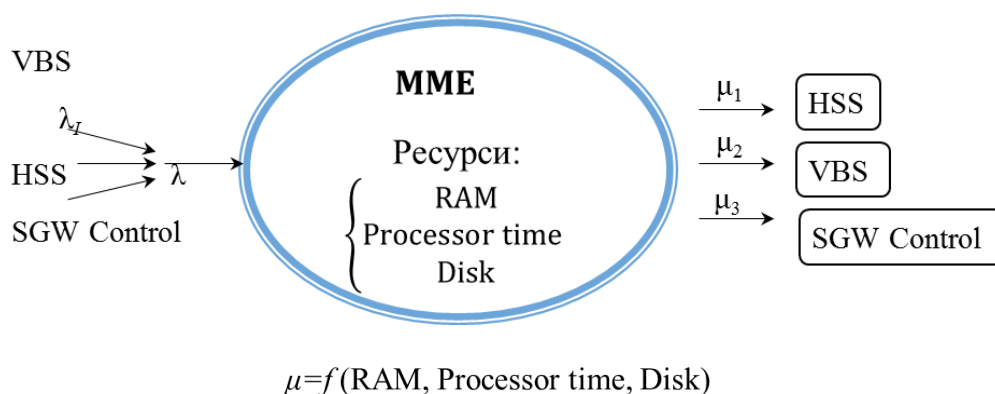


Рис. 4.5 Задача вибору ресурсів вузла обслуговування

Розрахунок ресурсів обслуговування необхідно проводити для кожного віртуального вузла обслуговування, на основі даних про інтенсивність надходження заявок. При розрахунку інтенсивність надходження покладається як величина з незмінюваним середнім значенням. Потік задач складається з різних запитів, визначених протоколами взаємодії вузла обслуговування. Нижче наведено набір запитів до підсистеми MME в процесі встановлення з'єднання.

Запити до MME:

1. S1AP: S1 Setup
2. Initial UE Attach
3. *Authn. Data Response*
4. S1AP: UL - Info Transport+ NAS: Authn. Response
5. S1AP: UL Transfer+NAS Security Mode Complete
6. *Insert Subscriber Data*

7. *Update Location Ack*
8. GTP: Create Session Response
9. S1AP: Initial Context Setup Response
10. S1AP: UL – NAS Transport
11. GTP: Modify Bearer Response

Вхідні дані:

- λ – середнє значення проміжків часу між моментами надходження замовлень,
- n – кількість заявок, які можуть одночасно обслуговуватися у віртуальному вузлі.
- v_k – кількість узагальненого обчислювального ресурсу, який використовується для обслуговування однієї заявки при одночасному обслуговуванні k заявок $k = \overline{1, n}$.
- a, b – верхня та нижня границя завантаженості системи обслуговування,
- R – задане експертами значення, який показує долю часу роботи системи без черги.
- V – кількість узагальненого ресурсу

Вихідні дані:

- μ – вектор інтенсивностей обслуговування заявок, елемент вектору μ_k відповідає інтенсивності обслуговування, якщо у системі перебуває k заявок, $k = \overline{1, n}$.

Основна задача: мінімізувати ресурси, які використовуються для обслуговування заявок у системі.

Була запропонована математична модель задачі вибору потужності вузла обслуговування, яка враховує ергодичний розподіл ймовірностей одночасного перебування у системі k -заявок. Також враховується залежність інтенсивності обслуговування від кількості заявок, які обслуговуються одночасно. Задача полягає у пошуку такого розподілу інтенсивностей

обслуговування, який би мінімізував загальну кількість ресурсів, що використовуються для обслуговування, утримував завантаженість системи у заданих межах, а також забезпечувала роботу системи без затримки із заданою ймовірністю.

Для вирішення задачі будуть застосовані формули для розрахунку ергодичного розподілу :

$$p_k = p_0 \prod_{i=1}^k \alpha_i, \quad (k = \overline{1, n});$$

$$p_{n+k} = p_0 \gamma_n^k \prod_{i=1}^n \alpha_i, \quad (k = 1, 2, \dots, m);$$

$$p_0 = \left(1 + \sum_{k=1}^n \prod_{i=1}^k \alpha_i + \frac{\gamma_n}{1 - \gamma_n} \prod_{i=1}^n \alpha_i \right)^{-1}$$

де $\alpha_k = \lambda / (k \mu_k)$, $\gamma_k = \lambda / (n \mu_k)$,

Цільова функція буде мати вигляд:

$$\sum_{k=1}^n k v_k p_k \rightarrow \min$$

$$\text{а саме: } \sum_{k=1}^n k v_k \frac{\prod_{i=1}^k \lambda / (i \mu_i)}{1 + \sum_{k=1}^n \prod_{i=1}^k \lambda / (i \mu_i) + \frac{\lambda / (n \mu_n)}{1 - \lambda / (n \mu_n)} \prod_{i=1}^n \lambda / (i \mu_i)} \rightarrow \min$$

функція є неперервною відносно змінних $\{\mu_1, \mu_k, \dots, \mu_n\}$,

При обмеженнях:

$$\frac{\lambda}{a} \geq \mu_i, \frac{\lambda}{b} \leq \mu_i$$

$$\sum_{k=1}^n \frac{\prod_{i=1}^k \lambda / (i \mu_i)}{1 + \sum_{k=1}^n \prod_{i=1}^k \lambda / (i \mu_i) + \frac{\lambda / (n \mu_n)}{1 - \lambda / (n \mu_n)} \prod_{i=1}^n \lambda / (i \mu_i)} = R$$

Задача вибору потужності вузла обслуговування є задачею нелінійного програмування в процесі моделювання використовувалася функція Matlab `fmincon`, яка здійснює ітераційний пошук, починаючи з початкового значення, який зупиняється при наближенні до оптимального значення із заданою похибкою.

4.4. Задачі розподілу ресурсів при обслуговуванні потоку гібридних телекомунікаційних сервісів на сервері онлайн тарифікації послуг запитів

Аналіз процесів обслуговування сервісів у хмарному середовищі показав, що є технічна можливість обслуговувати сервіси у віртуальному просторі з заданою кількістю обчислювальних ресурсів. В той же час загальна кількість ресурсів, яка орендується провайдером телекомунікаційних послуг, є обумовленою та обмеженою договором оренди. Різноманіття телекомунікаційних сервісів та різні вимоги до процесу обслуговування сервісу обумовлює необхідність вирішення наукової задачі розподілу обмеженої кількості обчислювальних ресурсів між віртуальним просторами, які обслуговують різні типи телекомунікаційних сервісів.

Для моделювання процесу обслуговування телекомунікаційних сервісів було розглянуто систему онлайн тарифікації оператора зв'язку. Оператор зв'язку для забезпечення потреб системи тарифікації орендує хмарний обчислювальний простір для забезпечення процесу онлайн тарифікації. При цьому, у хмарному середовищі розгортається програмне забезпечення, що обслуговує заявки на тарифікацію різних послуг.

Зростання конкуренції у сфері мобільного зв'язку привело до значного збільшення числа послуг, які пропонують абонентам мобільні оператори. Наслідком цього стало ускладнення процесу обслуговування абонентів на сервері тарифікації. Обслуговування абонентів на сервері тарифікації є необхідною складовою частиною процесу обслуговування абонентів та надання їм телекомунікаційних послуг. Оскільки кожна послуга є платною,

оператор зв'язку залежно від типу тарифного плану здійснює тарифікацію заявки або в перед наданням послуги в режимі реального часу або в режимі offline має перевірити наявність коштів на рахунку абонента, зробити перерахунок, тобто здійснити ряд стандартних операцій.

Значна кількість робіт присвячена забезпеченню ефективної тарифікації, при цьому розглядаються питання розробки правил PCC (Policy Charging Control), в яких задається політика обслуговування індивідуально для кожного типу сервісу залежно від тарифного плану [1, 2-4, 5]. Роботи [6, 7, 8] присвячені оптимізації білінгових систем і пов'язані з дисципліною обслуговування направлених на диференціацію вхідного потоку та створення нових правил послідовності тарифікації абонентів.

Однак, для операторів зв'язку достатньо гостро постає проблема перевищення часу обслуговування заявки саме на сервері тарифікації, оскільки не враховується принцип розподілу ресурсів технічних засобів, що у періоди пікових навантажень є критичними для якості обслуговування.

Одним із пріоритетних напрямків удосконалення системи обробки викликів в мережах мобільного зв'язку є забезпечення ефективного розподілу ресурсів на сервері оператора з урахуванням диференціації послуг за змістовною складовою.

4.4.1. Структура системи тарифікації. Сучасні системи обробки та тарифікації викликів в мобільних мережах зв'язку

Проблеми керування надання послуг пов'язані зі взаємодією процесів тарифікації та обробки викликів в сучасних телекомунікаційних мережах зв'язку, глибоко досліджені закордонними вченими Р. Kurtansky в праці [9], Н. Oumina [10], в книгах Camarillo [11], Khalid Al-Begain [12]. Теоретичні засади цих процесів розглянуті в роботах вітчизняних науковців Ільченка М.Ю. [13], Романова А.І.[14], Стеклова В.К. [15], Беркман Л.Н.[16], Вінницького В.П. [17], Зайченко О.Ю. [18].

Задачі керування якістю надання послуг в сучасних мобільних мережах зв'язку вирішуються на різних рівнях.

Перший тип задач – транспорт послуги, передбачає передачу послуги з гарантованою швидкістю передачі, затримкою, джиттером, коефіцієнтом втрат пакетів. Такі задачі вирішуються на 1-4 рівнях OSI, за допомогою протоколів маршрутизації [19, 20], методів пріоритетної обробки пакетів на проміжних вузлах і шлюзах – WFQ, CBQ, FIFO, LIFO, описані [21], алгоритмів розподілу ресурсу [22, 23, 24], гарантування якості при підвищенні утилізації ресурсів [25, 26], методів оцінки моніторингу та технічних показників якості мережі [27] в рамках систем мобільного зв'язку 2-го, 3-го, 4-го покоління [28, 29, 30 50-53].

Другий тип задач відноситься до керування якістю послуги на 5-му і вищих рівнях OSI. На цих рівнях вирішуються наступні задачі:

- забезпечення встановлення з'єднання з заданим коефіцієнтом відмов і затримки;
- узгодження рівня якості послуги з профілем абонента, станом його балансового рахунка та завантаженням мережі;
- керування якістю відповідно до запитів користувача;
- керування форматом представлення інформації;
- керування транспортним рівнем мережі для резервування необхідного ресурсу з метою гарантування якості подальшого обслуговування послуги; моніторинг параметрів якості, що впливають на тарифікацію послуг.

Задачі встановлення з'єднання розглядаються в роботах [31, 32, 33, 34]. У джерелі [55] досліджується затримка роботи кожної команди SIP протоколу [35, 36]. Автори статті виявили, що максимальне навантаження щодо обслуговування викликів в IMS системі припадає на функціональний блок обслуговування SIP викликів. У роботі [56] запропоновано модель продуктивності SIP протоколу, що базується на середній затримці очікування встановлення виклику і витраченій пропускній здатності для реалізації

протоколу. Наведено порівняння теоретичної моделі з вимірами і підтверджено адекватність моделі.

Зроблені висновки спонукають до більш детального вивчення варіантів реалізації з'єднання на базі SIP протоколу. Питання дослідження перевірки заданих параметрів якості при надходженні SIP запиту на з'єднання дослідженні в роботі [59]. У роботі показано важливість протоколу перевірки ресурсу мережі доступу для впровадження послуги – Resource Availability System (перевірки ресурсів системи). На основі результатів перевірки автори пропонують систему адаптації мультимедійного потоку, що пов'язана зі створенням адаптивного кодеку для передачі такого виду інформації.

Питання узгодження якості послуги з профілем абонента, станом його балансового рахунка і перевантаженням мережі розглядаються в роботах [55, 56, 37, 38, 39, 40]. У [55] описується нова архітектура керування якістю та тарифікацією, що дозволяє гнучко і незалежно надавати доступ до послуг згідно до сформованих правил.

У роботі [54] пропонується удосконалення системи за рахунок можливості введення динамічного керування послугами та тарифікацією шляхом переговорів. Користувач може задавати параметри тарифікації як діапазон вхідних параметрів для переговорного процесу. Наприклад, мінімальну та максимальну прийнятну ціну послуги. В залежності від заданих параметрів може бути впроваджена конфігурація, найближча до мінімальної або до максимальної ціни. Після встановлення результуючої конфігурації, параметри відправляють до блоку прийняття рішень щодо активації послуги. Конфігурація може бути змінена в переговорному режимі на будь-якому етапі надання послуги через зміну ресурсів мережі або зміну початкових вхідних параметрів. Реалізація такого підходу базується:

- по-перше, на модифікації обміну інформацією між користувачем, системою керування якістю та тарифікацією, мережею доступу;

- по-друге, на модифікації алгоритмів роботи блоків впровадження послуг і прийняття рішень (PCEF, PCRF,) що описані авторами на базі діаграм станів системи;

- по-третє, на створенні нового інтерфейсу взаємодії між блоком PCRF і підсистемою тарифікації OSC (система тарифікації), що пропонується для реалізації переговорного процесу між функціональними блоками в РСС. Також пропонується модифікація процедури встановлення сесії. Переваги такого підходу – в гнучкості присвоєння рівня якості обслуговування і адаптації відповідного тарифу.

Автори даної статі [54] не розкривають реалізації своєї пропозиції щодо керування якістю та тарифікацією. Недоліки такого підходу:

- відсутні правила формування залежності тарифу від якості та стану мережі, обмеження переговорного процесу;
- непрозорість отримання параметрів мережі, на основі яких будуються можливі конфігурації;
- не визначені механізми впливу користувача на параметри тарифікації, а саме: принцип, протоколи і програмні засоби.

Основними елементами керування пакетними послугами в сучасних системах є вузол GGSN (Gateway Service Node) – Обслуговуючий шлюзовий вузол, та SGSN (Serving GPRS Support Node) – Вузол обслуговування абонентів GPRS. Функції вузлів GGSN і SGSN детально описані [3]. Функціональна архітектура тарифікації пакетних послуг в мережах GSM зображена на Рис. 4.6.

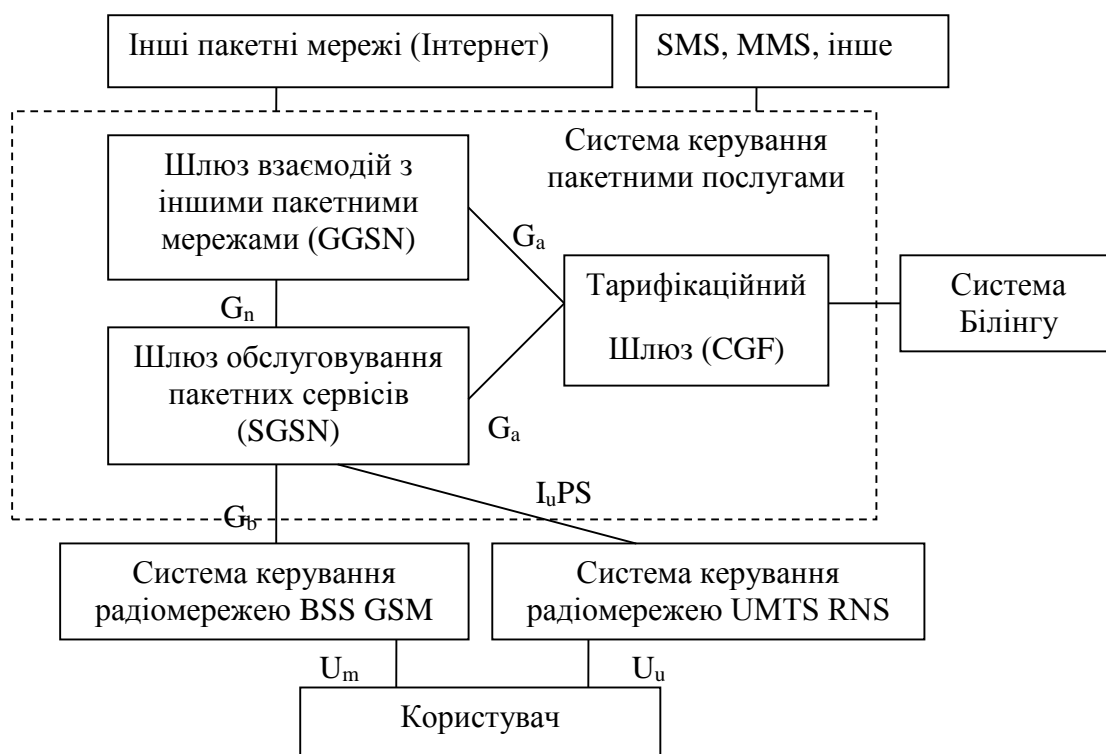


Рис. 4.6 Функціональна архітектура тарифікації пакетних сервісів GPRS/UMTS

Основною функцією шлюзу тарифікації є збір, зберігання, форматування та передача тарифікаційної інформації від вузлів GGSN і SGSN до білінгової системи. CGF підтримує два режими роботи. Перший режим – централізований, що підтримує передачу інформації від GGSN і SGSN в один елемент. Другий режим – децентралізований, в даному випадку функція CGF вбудована в вузли від GGSN і SGSN.

Детальний опис інтерфейсів U_m , U_u , G_b , G_n , G_a , I_uPS наведений в джерелі [41]. Інтерфейс G_a є інструментом передачі інформації CDR (charging data record) – тарифікаційний запис.

Наступна тарифікаційна інформація, визначена стандартом [42], що входить до CDR:

- кількість даних переданих і прийнятих користувачем, вимірюється в мегабайтах;
- протокол і якість обслуговування QoS в кожному з напрямків передачі інформації;
- тривалість передачі інформації;

- інформація щодо адреси відправника і адресату інформації;
- місце знаходження (номер БС) під час користування послугою;
- інша інформація.

При тарифікації CGF повинна враховувати зміну тарифу при наступних [43], подіях:

- зміна якості обслуговування,
- зміна часу дії тарифу;
- досягнення границі кількості переданої інформації, при котрій настає зміна тарифу.

Для передачі необхідної інформації між вузлами CGF і GGSN/SGSN використовують протокол GTP' – протокол, спеціально розроблений для передачі тарифікаційної інформації, при передачі GPRS трафіку за протоколом тунелювання GPRS Tunnelling Protocol [44].

Основним недоліками системи передачі пакетних сервісів GPRS є:

- по-перше, невелика швидкість передачі, максимально теоретична на абонента 172 кбіт/с;
- по-друге, існуючі механізми гарантування якості [45] не можуть гарантувати належних значень параметрів якості обслуговування мультимедійних послуг. Однією з найголовніших причин є те, що клас трафіку GPRS є менш пріоритетним порівняно з голосовими каналами. Тому неможливо гарантувати якість мультимедійних послуг, що надаються, наприклад, в Інтернеті. Додатково необхідно відмітити, що постійний ріст GPRS каналів в діапазоні 900 МГц або 1800 МГц не можливий в зв'язку з обмеженістю смуги.

Основні переваги реалізації сервісів в архітектурі системи тарифікації пакетних послуг GPRS [46]:

- точна настройка кожного сервісу окремо (SMS, MMS, доступ до Інтернет тощо), необхідних параметрів для тарифікації та інтерфейсів;

- порівняно просте керування тарифікацією. Тарифікація кожного сервісу – це окремий рівень (Рис. 4.7);

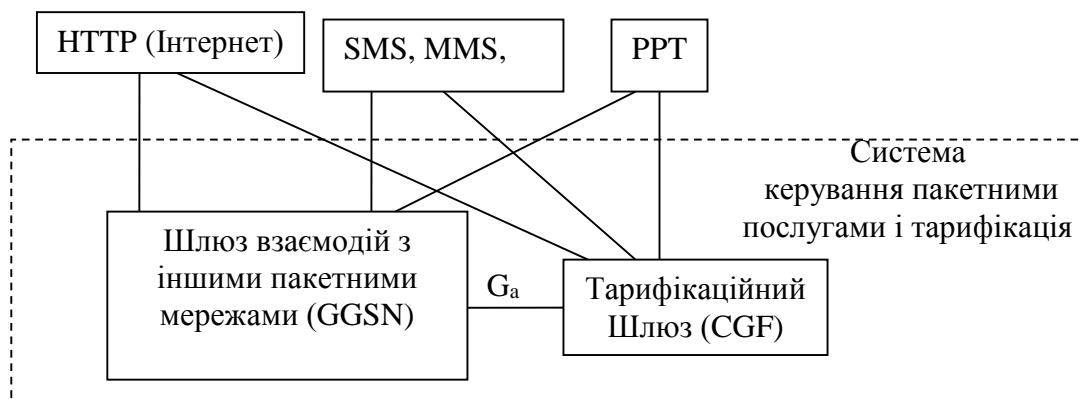


Рис. 4.7 Існуюча архітектура тарифікації пакетних сервісів в системі керування послугами GPRS/UMTS

Основними недоліками реалізації існуючої архітектури тарифікації пакетних сервісів в системі керування пакетними послугами GPRS/UMTS сервісів в архітектурі системи тарифікації пакетних послуг GPRS/UMTS [47] є:

- обмежена масштабованість системи. Для впровадження кожної
- послуги необхідна складна інтеграція в тарифікаційну систему;
- складна тарифікації при використанні одночасно багатьох послуг;
- неврахування якості при тарифікації пакетних послуг, що є загальним недоліком системи;
- тривалий час для введення послуги в експлуатацію;
- не враховує частотний ресурс при тарифікації послуг в мережах 4- го покоління з ОЧД.

Існуючі недоліки тарифікації пакетних послуг в мережах операторів GPRS/UMTS а саме: немасштабованість, неврахування якості при обслуговуванні та тарифікації [67-69], складність надання персоналізованих сервісів та їх тарифікації, – є одним з ключових факторів, що мотивували до розробки нових систем керування обробкою викликів в мобільних мережах зв'язку на базі протоколу IP.

Така система, запропонована в рамках проекту співробітництва 3GPP [74], отримала назву IMS система керування послугами на базі IP протоколу. Основними ініціаторами створення системи стали компанії Ericsson, Alcatel-Lucent, Huawei, Nokia-Simens, інші.

4.4.2. Процес обробки викликів в мобільних мережах зв'язку

Алгоритм методу обробки викликів послуг в мобільних мережах 4-го покоління представлений в Додатку Д. Передумови:

- термінал користувача зареєстрований в мережі оператора;
- запит передається після процедури ініціації з'єднання RRC і NAS [48] для бездротової мережі доступу;
- сервер і користувач знаходяться в «домашній» мережі.

Опис методу по крокам.

1. Блок P-CSCF отримує SIP запит на встановлення з'єднання, що включає опис програми прикладного рівня в тілі протоколу SDP[87]. P-CSCF присвоює сесії унікальний ICID, для визначення сесії, в тому числі тарифікації сесії, пересилає запит на блок-функцію S-CSCF.

2. S-CSCF приймає SIP запит, аналізує адресу прикладного інтерфейсу AF, наплавляє запит на API (інтерфейс програмування прикладних програмних компонентів)

3. AF приймає та аналізує SDP тіло, формує вимоги обслуговування послуги в мережі – швидкість, рівень якості, кодек. AF пересилає данні до PCRF

4. PCRF отримує від AF через Rx ідентифікатор прикладного програмного компонента з детальним описом вимог щодо якості і швидкості передачі, в тому числі тарифікаційний ключ, що визначає модель тарифікації: за часом, за Мбіт/с, інше.

5. PCRF створює правила обслуговування, включаючи тарифікаційні параметри, що мають відстежуватись для даного виду послуги. PCRF пересилає правила для впровадження PCEF.

6. PCEF отримує правила та впроваджує їх. PCEF ініціює процедуру перевірки або резервування каналу між абонентом і сервером послуг відповідно заданих PCRF правилами.

7. Якщо PCEF процедура, описана в п. 6, відбулася успішно, то автоматично відкривається тарифікаційний запис CDR, що зберігається у базі даних блоку PCEF.

8. Початок передачі послуги Користувач – Сервер на основі IP протоколу. Послуга повинна мати відповідний API інтерфейс і підтримувати протокол Diameter.

9. Під час передачі генеруються проміжні тарифікаційні записи CDR на основі відповіді на запит команди протоколу Diameter “UPDATE” до вузла PCEF.

10. Кінець передачі. SIP повідомлення про закінчення сеансу отримали P-CSCF, S-CSCF, AF, користувач.

11. Закривається тарифікаційний запис.

12. Відбувається кореляція записів, що отримана від елементів учасників з'єднання, в основному від PCEF. Здійснюється передача в CDR, в систему білінга.

Аналіз методу показує, що однією з основних процедур є виконання правил передачі (5,6), встановлення перевірки каналу за заданими параметрами і тарифікація, що відбувається в функціональних блоках PCEF, PCRF.

Існує два варіанти керування правилами, в разі встановлення з'єднання, для вхідних і вихідних викликів [49, 50]:

- примусове (push) керування політиками запитів;
- керування правилами «по запиту» користувача.

На Рис. 4.8 показана діаграма послідовності примусового керування політиками запитів [73].

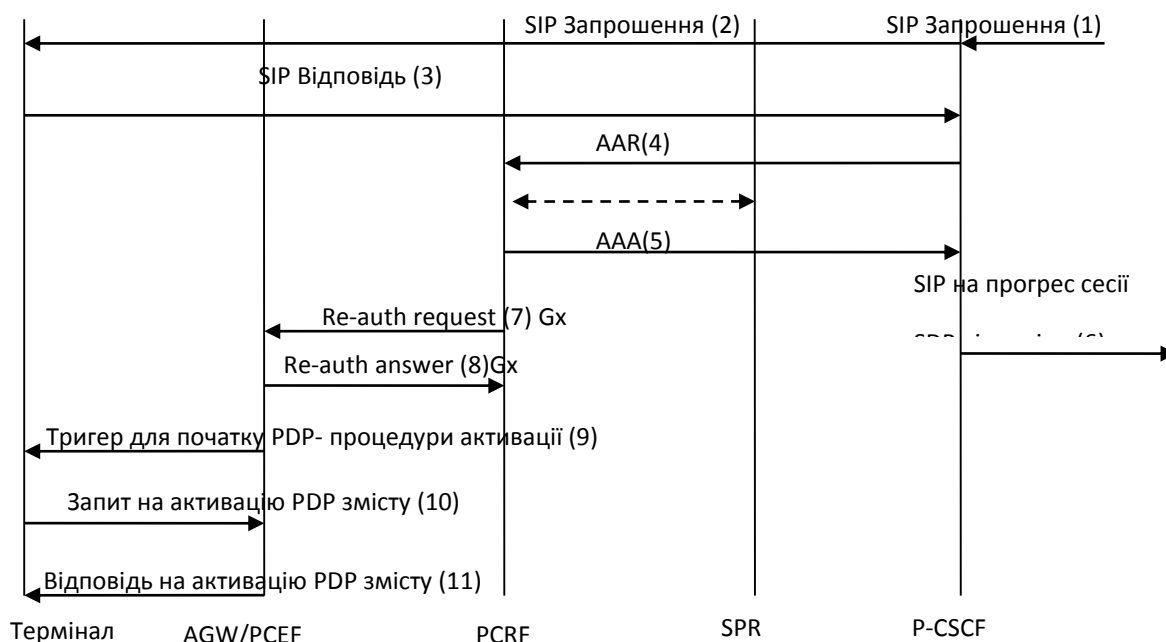


Рис. 4.8 Метод примусового керування правилами обробки вхідних запитів архітектури PCC

Метод керування обробкою викликів описаний нижче по крокам.

Блок P-CSCF отримує “SIP запрошення” на встановлення сесії з SDP описом сесії.

1. P-CSCF направляє запит відповідному терміналу на встановлення сесії.

2. Термінал відповідає на запит за допомогою повідомлення “SIP Відповідь”. Повідомлення містить згоду або відмову на встановлення з’єднання.

3. P-CSCF направляє за допомогою Diameter протоколу запит AAR (Authorization-Authentication(AA)-Request) - запит аутентифікації користувача і авторизації послуги та ресурсу для неї в PCRF. Запит включає опис медіа компоненту та суб-компоненту. Процедура перевіряє можливість надання послуги (через взаємодію з SPR) з необхідними характеристиками в мережі доступу.

4. PCRF відповідає на запит повідомленням “AA-Answer” – «успішно» або «не успішно». Під час цієї процедури будуються правила тарифікації і обслуговування послуги користувача.

5. P-CSCF посилає повідомлення – “SDP відповідь” щодо прогресу сесії на сервер ініціації виклику.

6. PCRF посилає запит в PCEF на виконання політики обслуговування та тарифікації викликів.

7. PCEF, через взаємодію з AGW, перевіряє можливість виконання політики обслуговування викликів, посилає відповідь до PCRF – “успішно” або “не успішно”.

8. PCEF посилає активаційний тригерний запит на термінал через PDP протокол (packet data protocol - пакетний протокол передачі даних та команд до користувача).

9. Термінал користувача посилає запит PDP з підтвердженням готовності початку передачі.

10. Шлюз відповідає – підтверджує активацію PDP запита.

11. Розглянуті механізми тарифікації та керування політиками вхідних запитів показують зв'язок між цими процесами.

Виходячи з аналізу сучасних систем тарифікації, слід відмітити особливості нових систем тарифікації, в тому числі на базі платформи IMS [72, 73]. Ці системи є конвергентними, можуть проводити тарифікацію послуг в будь-яких мережах доступу. Також нові системи тарифікації побудовані на основі функціональних архітектур, що включають функції керування тарифікацією і якістю передачі. Такі системи тарифікації враховують параметри, що детально описують послуги та субпослуги, використані кодеки при наданні послуги, якість обслуговування, період передачі, швидкість передачі, кількість переданої інформації.

Аналіз параметрів даних систем [72, 73, 51, 52] показав, що не враховані величини частотного ресурсу при тарифікації і наданні послуг в системах LTE, що базуються на ортогональному частотному доступі.

4.4.3. Обробка викликів на сервері тарифікації

Мобільна передача даних сьогодні є ключовим драйвером розвитку всієї стільникового галузі. Якщо обсяг голосового трафіка в мережах практично не росте, то трафік передачі даних, навпаки, експоненціально збільшується щорічно, і в найближчі роки ця тенденція збережеться.

Вся індустрія телекомунікацій розвивається з урахуванням цих тенденцій. Все більше інформаційних ресурсів переходять до подачі матеріалу у формі відеороликів, а не тексту з зображеннями. Все більше сервісів переїжджають в "хмару", а значить, вимагають гарантованої якості інтернет-з'єднання.

Велика кількість робіт та підходів присвячена контролю якості обслуговування абонентів. Розробляються методи забезпечення показників якості в процесі передачі інформації на різних етапах обслуговування [9, 10, 5]. Всі ці тенденції призводять до ускладнення процесу обслуговування викликів, тому сучасні системи обслуговування мають певні недоліки:

1. Неврахування важливих технічних та тарифікаційних складових послуг призводить до нераціонального використання ресурсів сервера.

2. Недостатньо ефективне керування навантаженням при збільшенні пропускної здатності системи обслуговування абонентів, ємностей каналів та серверних ресурсів, що забезпечують функцію тарифікації та обліку наданих послуг.

3. Розв'язок задачі керування в білінгових системах, який не задовольняє всім вимогам, може негативно впливати на якість послуги мобільної передачі даних.

Розподіл трафіка за типами дозволить більш раціонально використовувати ресурси сервера. Зараз все більше операторів переходять від динамічного до централізованого розподілу ресурсів.

Ідея розмежування трафіка за типами є перспективною для операторів, вона частково реалізована, але запропонована модель відрізняється тим, що

враховується не лише економічна ефективність від надання послуги, але й статистичні дані про навантаження на сервер у вибраний період часу, що дозволяє врахувати розподіл сервісів у вхідному потоці та забезпечити ефективну обробку у різні періоду часу.

Процес тарифікації викликів, які надходять до оператора мобільного зв'язку, відбувається на сервері OCS (Online Charging System). Це сервер кредитного контролю в режимі реального часу. Здійснює тарифікацію послуг, контролює баланс абонента, виконує обробку інформації про надходження та списання коштів на рахунку абонента.

Схему обслуговування заявок на сервері тарифікації показано на Рис. 4.9. Заявки надходять на сервер за різними протоколами (Diameter, CAP2, MAP), у відповідних модулях EDP (Enhanced Diameter Proxy) та RES і розподіляються між пулом адаптерів відповідних протоколів. Після чого декодуються у відповідних адаптерах CAP Protocol Adapter та Diameter Protocol Adapter та приводяться до єдиного виду.

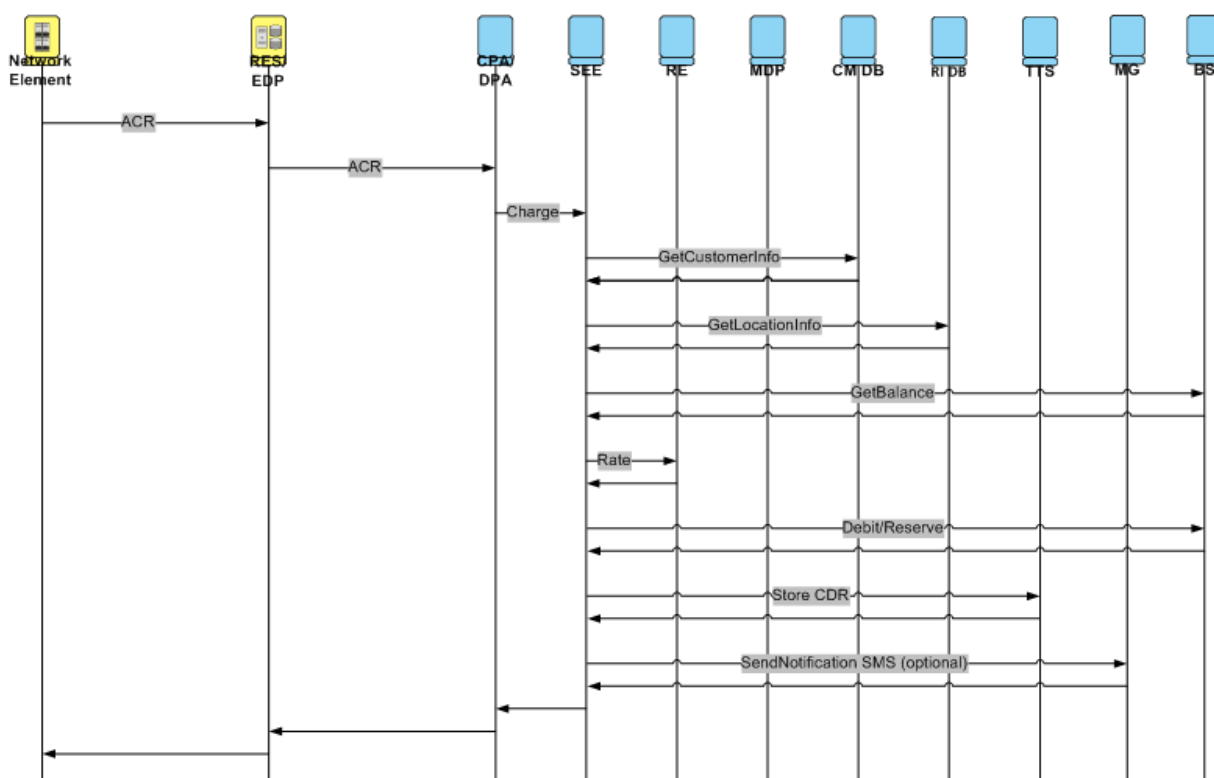


Рис. 4.9 Загальна схема online тарифікації послуг на сервері тарифікації

Розкодовану справу модуль передає на сервер бізнес логіки SEE через модуль маршрутизації BUS. Сервер бізнес логіки SEE являється ядром системи тарифікації та забезпечує середовище для виконання послідовності операцій, які передбачає процес обслуговування заявок.

Послідовність операцій, які повинні бути виконані для успішного обслуговування заявки, включає в себе наступні дії:

1. Вилучення інформації про абонента. Для цього модуль SEE звертається до бази даних керування абонентами CM DB.
2. Вилучення інформації про місце розташування абонента. Для цього модуль SEE звертається до бази даних, що зберігає структуру мережі RI DB.
3. Вилучення інформації про стан рахунку абонента. Для цього модуль SEE звертається до бази даних рахунків абонентів (Balance Storage).
4. Здійснюється розрахунок вартості послуги на основі тарифної книги та поточної тарифної моделі абонента. Для цього модуль SEE звертається до модулю розрахунків (RE).
5. З рахунку абонента знімається плата за послугу. Для здійснення операцій дебету та резервації коштів модуль SEE звертається до бази даних рахунків абонентів (Balance Storage).
6. Формується звіт про надання послуги CDR. Для цього модуль SEE звертається до програмного модулю формування CDR (Toll Ticket Server).
7. Якщо це передбачено послугою, відправляється повідомлення абоненту про результати наданих послуг. Для цього модуль SEE звертається до модулю для посилки повідомлень абоненту (MG).
8. Процес тарифікації завершено.

Як видно процес тарифікації є багатостадійним, при цьому операції, які послідовно виконуються у ядрі бізнес логіки SEE із залученням різних підсистем, різноманітні, відповідно потребують різної кількості оперативної пам'яті, процесорного часу та дискового простору.

4.4.4. Задача розподілу ресурсів сервера

Кожна послуга, що надається оператором мобільного зв'язку характеризується певним типом заявки, що надходить на сервер. Заявкою будемо вважати запит абонента на виконання певної послуги. При цьому у процесі обслуговування заявки виникає кілька важливих проблем. По-перше, на сервер одночасно надходить велика кількість неоднорідних заявок, які потребують негайної обробки. По-друге, вхідний потік заявок нерівномірний. Третя проблема пов'язана із неоднорідністю використання ресурсів серверу при забезпечуванні обслуговування кожної заявки. Серед основних ресурсів сервера можна виділити такі, як оперативна пам'ять, процесорний час, об'єм постійної пам'яті на дисках.

Обслуговування заявки на сервері складається з виконання деякої послідовності операцій, які можна розділити на логічно завершені етапи. Далі називатимемо ці етапи функціональними блоками. Для обслуговування абонентських заявок у функціональних блоках застосовують вищезазначені ресурси. Успішне проходження всіх функціональних блоків у заданій послідовності забезпечує успішне обслуговування заявки на сервері. Час обслуговування заявки на сервері є обмеженим, тому якщо заявка перебуває в системі довше заданого часу, вона знімається з обслуговування, і клієнту повідомляється, що мережа зайнята. Таким чином, щоб звести втрати заявок до мінімуму і максимізувати економічну ефективність обслуговування викликів, необхідно забезпечити оптимальний розподіл ресурсів сервера між різними типами заявок, що надходять до оператора.

Вхідними даними є об'єм кожного виду ресурсу, необхідний для обслуговування певного типу заявок у конкретному функціональному блоці, кількість прибутку від усіх видів заявок і максимальна кількість ресурсів, яку може надати сервер. Необхідно визначити кількість ресурсів, яку потрібно виділити для обробки певного типу заявок.

Нехай k_i – кількість заявок i -го типу; R_g – g -й тип ресурсу ($g=\overline{1, G}$, де G – загальна кількість типів ресурсів на сервері); v_{ij}^{Rg} – кількість певного ресурсу, необхідна для обслуговування однієї заявки i -го типу у j -му функціональному блоці; S_i – кількість прибутку, який можна отримати від однієї заявки i -го типу.

Тоді кількість ресурсів, необхідних для проходження заявки i -го типу через всі функціональні блоки буде рівна:

$$\sum_j v_{ij}^{Rg}$$

Загальна кількість прибутку, отриманого від обслуговування усіх заявок i -го типу дорівнює:

$$S = \sum_i k_i S_i$$

Загальна кількість ресурсу, яку займають всі заявки по всіх функціональних блоках дорівнює:

$$v^{Rg} = \sum_i k_i \left(\sum_j v_{ij}^{Rg} \right) \quad (40.1)$$

Оскільки необхідно максимізувати прибуток, тому цільова функція буде наступною:

$$\sum_i k_i S_i \rightarrow \max$$

Обмеження задачі полягає в тому, щоб кількість ресурсів, необхідна для обслуговування знайденої кількості заявок не перевищувала максимальну кількість ресурсів, яку може надати сервер:

$$\sum_i k_i \left(\sum_j v_{ij}^{Rg} \right) \leq V_{Rg}$$

Якщо розглядати окремо кожен тип ресурсів, отримаємо наступну систему:

$$\begin{cases} \sum_i k_i (\sum_j v_{ij}^{R1}) \leq V_{R1} \\ \dots \\ \sum_i k_i (\sum_j v_{ij}^{RG}) \leq V_{RG} \end{cases}$$

Де V_{R1} – кількість ресурсу першого типу, яку може надати сервер;

V_{RG} – кількість ресурсу G-го типу, яку може надати сервер.

Кількість рівнянь системи відповідає кількості типів ресурсів на сервері. Результатом вирішення задачі на умовний екстремум буде кількість заявок кожного типу, яку необхідно подати на вхід сервера, щоб забезпечити максимізацію прибутку від обслуговування усіх заявок. Для визначення розподілу ресурсів між різними типами заявок можна скористатись формулою (40.1).

При вирішенні задачі розподілу ресурсів необхідно також враховувати статистичні дані про надходження заявок на сервер у різні періоди часу.

На Рис. 4.10 зображена статистика надходження заявок на сервер протягом доби. Оскільки співвідношення між кількістю різних заявок є відомою величиною, нескладно накласти додаткові обмеження на кількості заявок, які будуть отримані на виході задачі.

Додаткові обмеження задачі, отримані в наслідок аналізу добової статистики – це відношення середньостатистичної кількості заявок різних типів сервісів:

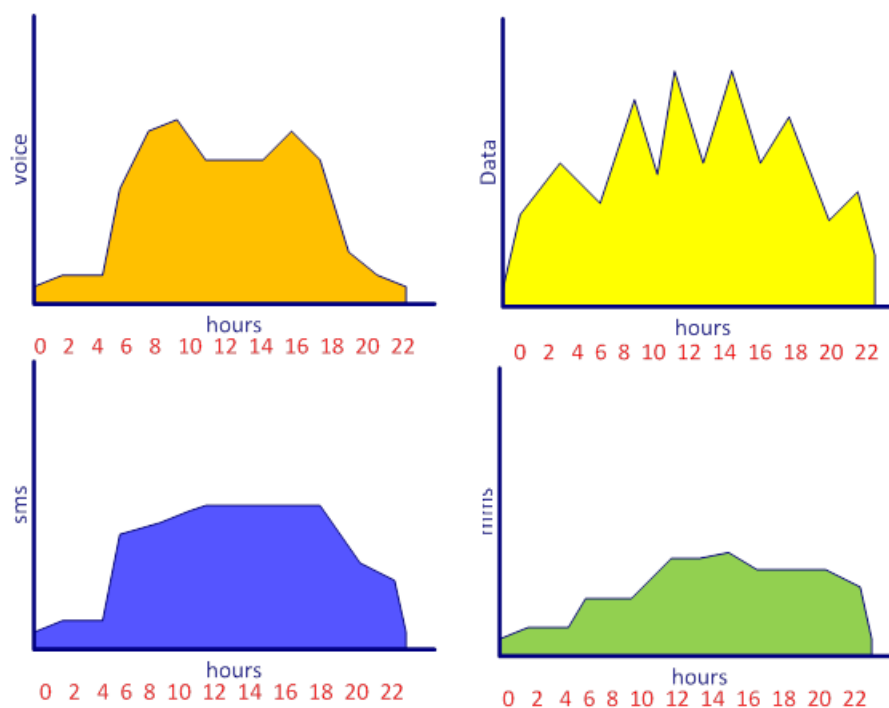


Рис. 4.10 Статистика надходження заявок протягом доби

$$a_{ij} \leq \frac{k_i}{k_j} \leq b_{ij}, \forall i, j = \overline{1, m}$$

де k_i та k_j – шукані кількості заявок на тарифікацію, що надійшли від i -го та j -го типів сервісів відповідно; a_{ij} та b_{ij} – числові границі розраховані на основі статистичних даних про середньогодинну кількість заявок на тарифікацію, що надходять до системи; m – кількість сервісів, для яких здійснюється тарифікація.

4.4.5. Оцінка ефективності методу розподілу ресурсів

Імітаційна модель, створена для реалізації запропонованого методу показала, що при фіксованому розподілі доступних технічних ресурсів між різними типами заявок, економічна ефективність роботи системи покращена на 5%, крім того, кількість втрачених заявок через перевищення допустимого часу обслуговування, а також внаслідок роботи алгоритмів раннього попередження перевантажень, які традиційно застосовують за для уникнення перевантажень скоротилася в середньому на 1%.

Було проведено два експерименти роботи імітаційної моделі для трьох типів заявок на тарифікацію: текстові повідомлення (sms), мультимедійні повідомлення (mms) та дзвінки. В першому на обслуговування виділявся обмежений технічний ресурс, який спільно використовувався для обслуговування заявок на тарифікацію, в процесі обслуговування застосовувався алгоритм RAD (Random early detection), який передбачав відкидання заявок в разі перевантаження серверу, крім того заявки втрачалися якщо час перебування їх у системі перевищував максимально допустиме значення.

В другому експерименті технічні ресурси розділялися відповідно до запропонованого методу. На основі статистичних даних оператора мобільного зв'язку про кількість заявок, що надходить на сервер протягом дня та кількість прибутку від обслуговування заявки певного типу, був розрахований об'єм ресурсу, що виділяється для обслуговування кожного з трьох типів заявок. Наприклад, кількість постійної пам'яті складає 1,25Мбайт для голосу, 9,15Мбайт для sms і 17,5 Мбайт для mms- повідомлень. В ході експериментів була імітована робота системи тарифікації з вхідним мультисервісним потоком максимально наближеним до реального.

Не зважаючи на те, що в сумі кількість втрачених заявок суттєво не змінилася, однак спостерігався перерозподіл кількості заявок між різними типами, які були втрачені, за рахунок чого спостерігалось зменшення розміру сумарного втраченого прибутку в другому експерименті.

Розрахунок проводився виходячи з того, що одночасно в середньому надходить по 10 тисяч заявок на тарифікацію, кожного з типів сервісу. В результаті в ході першого експерименту середні втрати заявок склали 2%, в тому числі втрати заявок на тарифікацію sms сервісів склали – 2,0%, mms сервісів – 2,55%, голосу – 1,45%. В ході другого експерименту середні втрати заявок склали 1,8% в тому числі втрати заявок на тарифікацію sms сервісів склали – 1,65%, mms сервісів – 1,25%, голосу – 2,5%.

В результаті експерименту було визначено, що для першого випадку кількість втрат складає: 870 у.о. для заявок 1 типу (дзвінки), 1800 у.о. для заявок 2 типу (sms) та 3825 у.о. для заявок 3 типу (mms). Теоретично можлива кількість отриманого прибутку при умові, що всі заявки, які надійшли на сервер будуть оброблені становить 100000 у.о. Загальна кількість втрат складає 6495 у.о., що становить 6,5% від можливого прибутку. Для випадку з застосуванням запропонованого методу були отримані наступні дані про кількість втраченого прибутку: 1500 у.о. для заявок 1 типу (дзвінки), 1485 у.о. для заявок 2 типу (sms) та 1875 у.о. для заявок 3 типу (mms). Загальна кількість втрат складає 4860 у.о., що становить 4,86% від можливого прибутку.

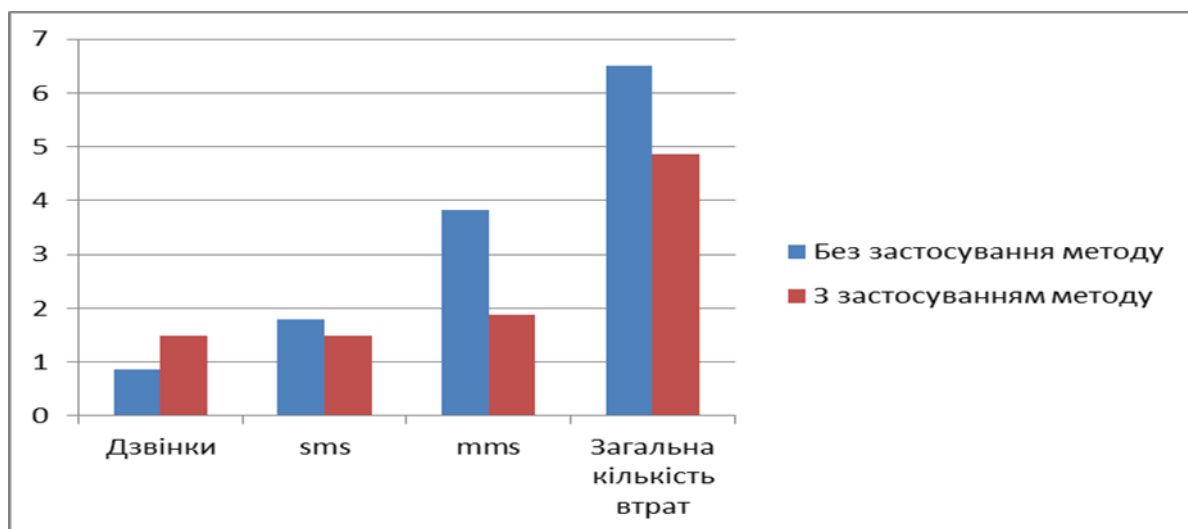


Рис. 4.11 Порівняння розміру втраченого прибутку у відсотках

На Рис. 4.11 представлено порівняння розміру втраченого прибутку у відсотках для кожного типу заявок та загальної кількості втрат до та після застосування методу розподілу ресурсів між різними типами заявок. Результат експерименту показав, що загальна кількість втраченого прибутку зменшилася на 1,64%.

Запропонований метод розподілу ресурсів системи для забезпечення ефективної обробки заявок дозволяє здійснювати контроль за якістю надання послуг, враховує необхідну кількість ресурсів для обслуговування однієї заявки, що дозволяє виділяти ресурси пропорційно вимогам сервісу, враховує

статистичні дані про кількості заявок різних типів сервісів, які надходять у заданих інтервалах часу, що дозволяє налаштовувати розподіл технічних ресурсів, що обслуговують заявки, відповідно до складу вхідного навантаження, а також забезпечити максимізацію економічної ефективності процесу надання послуг.

4.5. Методи розподілу технічних ресурсів для гібридного телекомунікаційного потоку

Аналіз процесів обслуговування сервісів у хмарному середовищі показав, що є технічна можливість обслуговувати сервіси у віртуальному просторі з заданою кількістю обчислювальних ресурсів. В той же час загальна кількість ресурсів, яка орендується провайдером телекомунікаційних послуг, є обумовленою та обмеженою договором оренди. Різноманіття телекомунікаційних сервісів та різні вимоги до процесу обслуговування сервісу обумовлює необхідність вирішення наукової задачі розподілу обмеженої кількості обчислювальних ресурсів між віртуальним просторами, які обслуговують різні типи телекомунікаційних сервісів.

Для моделювання процесу обслуговування телекомунікаційних сервісів було розглянуто систему онлайн тарифікації оператора зв'язку. Оператор зв'язку для забезпечення потреб системи тарифікації орендує хмарний обчислювальний простір для забезпечення процесу онлайн тарифікації. При цьому, у хмарному середовищі розгортається програмне забезпечення, що обслуговує заявки на тарифікацію різних послуг.

Кожна послуга, що надається оператором мобільного зв'язку характеризується певним типом заявки, що надходить на сервер. Заявкою будемо вважати запит абонента на виконання певної послуги. При цьому у процесі обслуговування заявки виникає кілька важливих проблем. По-перше, на сервер одночасно надходить велика кількість неоднорідних заявок, які потребують негайної обробки. По-друге, вхідний потік заявок нерівномірний. Третя проблема пов'язана із неоднорідністю використання ресурсів серверу

при забезпечуванні обслуговування кожної заявки. Серед основних ресурсів сервера можна виділити такі, як оперативна пам'ять, процесорний час, об'єм постійної пам'яті на дисках.

Обслуговування заявки на сервері складається з виконання деякої послідовності операцій, які можна розділити на логічно завершені етапи. Далі називатимемо ці етапи функціональними блоками. Для обслуговування абонентських заявок у функціональних блоках застосовують вищезазначені ресурси. Успішне проходження всіх функціональних блоків у заданій послідовності забезпечує успішне обслуговування заявки на сервері. Час обслуговування заявки на сервері є обмеженим, тому якщо заявка перебуває в системі довше заданого часу, вона знімається з обслуговування, і клієнту повідомляється, що мережа зайнята. Таким чином, щоб звести втрати заявок до мінімуму і максимізувати економічну ефективність обслуговування викликів, необхідно забезпечити оптимальний розподіл ресурсів сервера між різними типами заявок, що надходять до оператора.

Для обслуговування абонентських заявок у функціональних блоках застосовують ресурси, такі як оперативна пам'ять, процесорний час, мережевий ресурс (зайнятість каналу сигнальним трафіком), об'єм постійної пам'яті на дисках. Обслуговування кожної заявки у заданому функціональному блоці, потребує заданого об'єму ресурсу. Відомо протягом якого часу використовується кожний з ресурсів при обслуговуванні у заданому функціональному блоці. Якщо ресурс зайнятий, тоді заявка очікує звільнення ресурсу. Таким чином, виникають затримки в обслуговуванні, які ведуть за собою втрату успішно обслугованих заявок. Саме тому система моніторингу використання ресурсів різними типами сервісу побудована з урахуванням особливостей етапів обслуговування заявок на сервері оператору мобільного зв'язку буде корисною при організації системи керування трафіком.

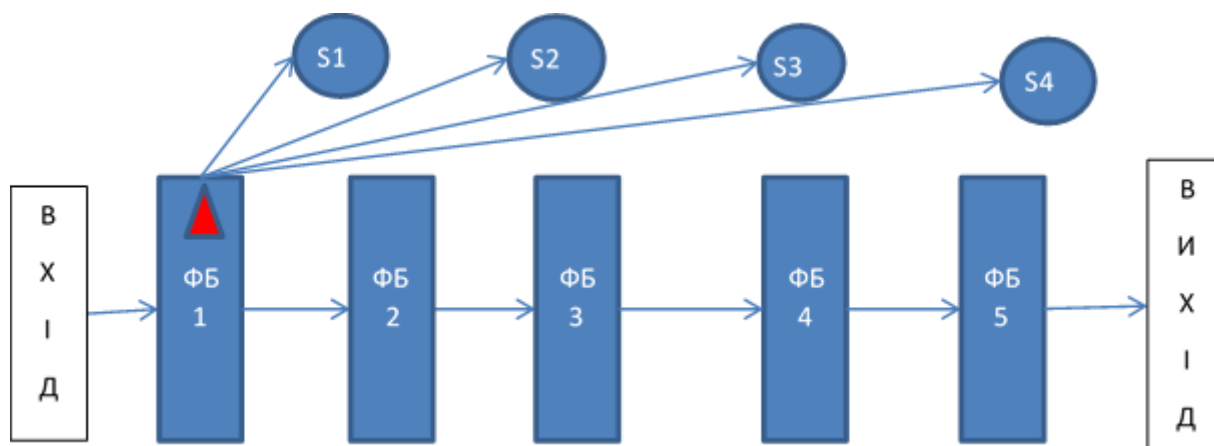


Рис. 4.12 Схема обслуговування заявки на сервері

На Рис. 4.12 показана схема обслуговування заявки (позначена трикутником), яка надійшла на обслуговування в систему, та обслуговується у першому функціональному блоці (ФБ1), відповідно обслуговування потребує використання ресурсів (стрілками показані запити до ресурсів). Обслуговування заявки буде успішним, якщо вона пройде від «Входу» до «Виходу» за допустимий час.

4.5.1. Задача визначення розподілу кількості заявок які у поточний момент часу знаходяться на обслуговуванні в системі.

Постановка задачі. Заявки на обслуговування надходять до системи за заданим законом. Процес обслуговування однієї заявки включає в себе перебування (обслуговування) заявки в одному з n функціональних блоків, для обслуговування використовується G типів ресурсів. Нехай на обслуговування поступають заявки від m типів сервісів. Відома статистика часу перебування заявки i -го типу сервісу ($i = \overline{1, m}$) в j -му функціональному блоці ($j = \overline{1, n}$). Необхідно знайти розподіл заявок між функціональними блоками та за типами сервісів, які обслуговуються в системі у поточний момент часу.

Визначимо матрицю $T = \{ t_{ij} \}$, кожний елемент якої відповідає математичному очікуванню часу перебування заявки i -го типу ($i = \overline{1, m}$) в j -

му функціональному блоці ($j = \overline{1, n}$), що розраховується за допомогою методу моментів, метод розрахунку винесений у окремий підпункт статті.

Визначимо матрицю $V^{sg} = \{v_{ij}^{sg}\}$, кожний елемент v_{ij}^{sg} якої відповідає об'єму ресурсу g -го типу який використовується при обслуговуванні заявки i -го типу ($i = \overline{1, m}$) в j -му функціональному блоці ($j = \overline{1, n}$). Тоді сумарна кількість ресурсу g -го типу ($g = \overline{1, G}$), що займається i -м сервісом у поточний момент часу визначається за формулою

$$v_i^{sg} = \sum_{j=1}^n k_{ij} v_{ij}^{sg} \quad (4.2)$$

де k_{ij} – кількість заявок i -го типу, що обслуговуються у поточний момент часу у j -му функціональному блоці;

v_i^{sg} - об'єм ресурсу g -го типу, який займається заявками i -го типу сервісу у поточний момент часу.

Розглянемо *приклад 1*. Нехай $m=1$, $n=4$. Тобто система обслуговує 1 тип заявок у 4х функціональних блоках, для обслуговування потрібна тільки оперативна пам'ять ($G=1$, позначимо $s1$).

Нехай вже розраховано за Методом моментів математичне очікування часу перебування заявки у кожному функціональному блоці: $t_{11}=1$, $t_{12}=3$, $t_{13}=4$, $t_{14}=1$.

$$T = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 1 \end{pmatrix}$$

Тобто в середньому заявка проводить у системі 9 одиниць часу, позначимо його за T_{Σ} ($T_{\Sigma}=9$). Нехай протягом 9 одиниць часу до системи надійшло $K=100$ заявок, нехай вхідний потік описується рівномірним розподілом, тоді кількість заявок, яка попала у відповідний функціональний блок та обслуговується у поточний момент часу буде наступною:

$$k_{1j} = \frac{t_{1j}}{T_{\Sigma}} K,$$

де k_{1j} – кількість заявок першого типу у j -му функціональному блоці.

Нехай відомо, що для виконання операцій з інформаційним потоком 1-го типу у 1-му функціональному блоці потрібен об'єм оперативної пам'яті 1 одиниця ($v_{11}^{s1} = 1$), для виконання операцій з 1м інформаційним потоком у функціональному блоці 2 потрібен об'єм оперативної пам'яті 0 одиниць ($v_{12}^{s1} = 0$), відповідно: $v_{13}^{s1} = 4$, $v_{14}^{s1} = 2$. Тобто матриця V^{s1} має вигляд:

$$V^{s1} = \begin{pmatrix} 1 \\ 0 \\ 4 \\ 2 \end{pmatrix}$$

На Рис. 4.13 наведено використання ресурсу $s1$ протягом часу обслуговування однієї заявки 1го типу в усіх 4х функціональних блоках.

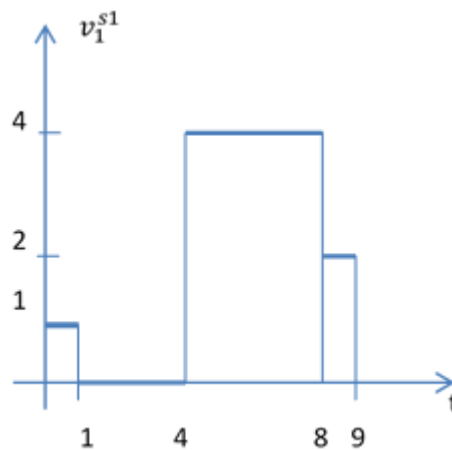


Рис. 4.13 Приклад використання ресурсу в часі

Оскільки за припущенням протягом дев'яти одиниць часу до системи рівномірно надходили 100 заявок, тоді можна говорити про наступний розподіл ресурсу $s1$ між заявками, які знаходяться на обслуговуванні. Об'єм ресурсу $s1$, який займають 100 заявок визначається за формулою: $v_1^{s1} =$

$$\sum_{j=1}^4 k_{1j} v_{1j}^{s1} = \sum_{j=1}^4 \frac{t_{1j}}{T_{\Sigma}} K v_{1j}^{s1}$$

$$v_1^{s1} = \frac{1}{9} * 100 * 1 + \frac{3}{9} * 100 * 0 + \frac{4}{9} * 100 * 4 + \frac{1}{9} * 100 * 2 = 211,11.$$

□

Припущення про рівномірний вхідний потік є суттєвим для моделі. В реальних системах вхідний потік заявок від сервісів може бути описаний за

допомогою Пуассонівського закону розподілу випадкових величин. Таким чином, задачу аналізу завантаженості ресурсів серверу треба вирішити у два підходи:

1. Провести аналіз використання ресурсів системи за умови, що відомо скільки заявок та якого типу сервісу надходило до системи у кожний період часу, що дозволяє побудувати систему моніторингу, яка визначить поточні проблеми щодо використання ресурсів системи.

2. Провести аналіз використання ресурсів системи протягом тривалого часу, для чого потрібно проаналізувати вибірки великого об'єму для визначення «вузьких місць» у системі. Така задача буде розглянута у подальших роботах.

4.5.2. Передумови створення системи керування вхідним потоком на вузол хмарного середовища

Як було показано в 3 розділі процес тарифікації є багатоетапним, при цьому операції, які послідовно виконуються у ядрі бізнес логіки SEE із залученням різних підсистем, різноманітні, відповідно потребують різної кількості оперативної пам'яті, процесорного часу та дискового простору. При вирішенні задачі керування вхідним потоком необхідно звернути увагу на час використання ресурсів. Необхідно врахувати як загальну кількість ресурсів, що обслуговує сервер в цілому, так і розділення ресурсів (методами віртуалізації), які з одного боку забезпечують ефективне обслуговування кожного етапу обробки, а з іншого є обмеженням, оскільки підсистема використовує лише ресурси їй відведені та не має доступу до інших ресурсів. Є також необхідність врахувати розподіл середнього часу виконання операцій.

Друга особливість полягає у тому, що кожний тип послуг, незважаючи на стандартність операцій, які виконуються в процесі тарифікації, потребує для здійснення розрахунків різної кількості ресурсів.

З точки зору необхідної процедури обслуговування всі сервіси можна розділити на три групи:

- Сесія тарифікації з резервацією (SCUR - Session Charging with Unit Reservation) оперативна пам'ять зайнята на всю тривалість сесії (може тривати до доби – наприклад, GPRS)
- Моментальна тарифікація події (IEC Immediate Event Charging) не зберігає стан свого виконання в пам'яті – виконується оцінка та списання грошових коштів в один момент (SMS).
- Тарифікація події з резервацією (ECUR - Event Charging with Unit Reservation) – оперативна пам'ять зайнята на період резервації (наприклад, час доставки контенту абоненту: відео, музика, MMS).

Таким чином, сервіси SCUR та ECUR виконуються у декілька етапів. Стан заявки або стан виклику зберігається на у підсистемі MDP (Memory DataBase Provider). MDP – це модуль для збереження поточного стану, що представляє собою програмно-апаратний комплекс, який забезпечує швидкий доступ до оперативної пам'яті (запис, зчитування, пошук).

На Рис. 4.12 було наведено загальну схему online тарифікації всі сім етапів одноразово проходять заявки групи сервісів IEC. При обслуговуванні сервісів SCUR та ECUR перший та другий етапи, що включають в себе вилучення інформації про абонента та про місце його розташування виконуються один раз після чого вся інформація про абонента та стан заявки-виклику зберігається у підсистемі MDP.

Ресурси, які потребує система при обслуговуванні 7 етапів залежать не тільки від групи сервісів але й від його типу. Відмінність у швидкості та ресурсозатратності операцій виникає при розрахунку вартості послуги та при вилученні інформації про стан заявки-виклику із системи MDP. Ці умови мають бути враховані при розрахунку плану керування вхідним потоком на сервер тарифікації.

Третя особливість полягає в тому, що одночасно надходить велика кількість заявок на тарифікацію різних типів ресурсів.

На сьогоднішній день оператори мобільного зв'язку надають послуги мільйонам абонентів, наприклад компанія «Київстар» обслуговує до 26 мільйонів абонентів. При цьому за рахунок привабливих пакетних умов все більше абонентів користуються послугами мобільного Інтернету, та дзвінків. За умови централізованого обслуговування система тарифікації одночасно обслуговує до одного мільйона абонентів, які замовляють або продовжують користуватися послугами. Для кожної заявки абонента ініціюється ланцюг операцій описаних вище. У часи найбільшого навантаження кількість абонентських заявок збільшується у декілька разів.

Четвертою особливістю є неоднорідність вхідного потоку заявок. Системи обслуговування абонентів прийнято розглядати як системи із пуассонівським вхідним потоком заявок. Основними особливостями якого є значна дисперсія кількості заявок, що надходять на тарифікацію. Для розподілу Пуассона дисперсія дорівнює математичному очікуванню. Тобто можливі сплески навантаження на короткий період часу, що є меншим за час обслуговування заявки на сервері тарифікації. Такі сплески призводять до тимчасового перевантаження серверу навіть в умовах не часу пік.

З точки зору реалізації логіки процесу обслуговування, роботу підсистем серверу оператора мобільного зв'язку можна представити як багаторівневу систему масового обслуговування, де керування потоком заявок здійснюється на двох рівнях.

Перший логічний рівень прикладних програмних компонентів. Тут заявки, які надійшли у систему, розділяються за типом сервісу, який вони представляють, обслуговування черг ведеться відповідно до схеми обслуговування розробленої для відповідного типу сервісу. Процес керування включає в себе формування черг за типом сервісу, застосування методів групи WRAD, а також інших схем керування, які враховують специфіку обслуговування сервісів [вставить ссылку про патент]. Таким

чином, система масового обслуговування першого рівня представляє собою заявки різних типів сервісів, які надходять на обслуговування до системи, будемо називати їх заявками першого рівня. Обслуговуючими пристроями у такій системі виступають ланцюги функціональних блоків, де здійснюється послідовне обслуговування заявок, кожний тип сервісу обслуговується у окремому ланцюзі.

Другий рівень – рівень технічної обробки. Схема обслуговування заявок за типом сервісу передбачає послідовне виконання операцій, для виконання яких потрібна задана кількість апаратних ресурсів, кожену операцію можна представити як заявку на обслуговування, будемо говорити про заявки другого рівня, де обслуговуючими пристроями виступають апаратні ресурси. Тут заявки другого рівня організовуються у черги до відповідних ресурсів. Політики використання ресурсів визначаються методами керування ресурсами обчислювальної системи. Архітектури розподілу ресурсів, організація обслуговування заявок другого рівня, суттєво впливають на швидкість обслуговування. Однак, така архітектура системи обробки заявок другого рівня є постійною, про її роботу можна судити по статистичним даним затримок в обслуговуванні заявок першого рівня.

Оскільки, вхідний потік заявок другого рівня однозначно визначається кількістю заявок першого рівня, що обслуговуються у системі. Тому система керування вхідним потоком заявок першого рівня, яка побудована з урахуванням статистики завантаженості ресурсів системи другого рівня, дозволить зменшити втрати заявок через затримки пов'язані з нестачею ресурсів.

Архітектура дворівневої системи керування представлена на Рис. 4.14

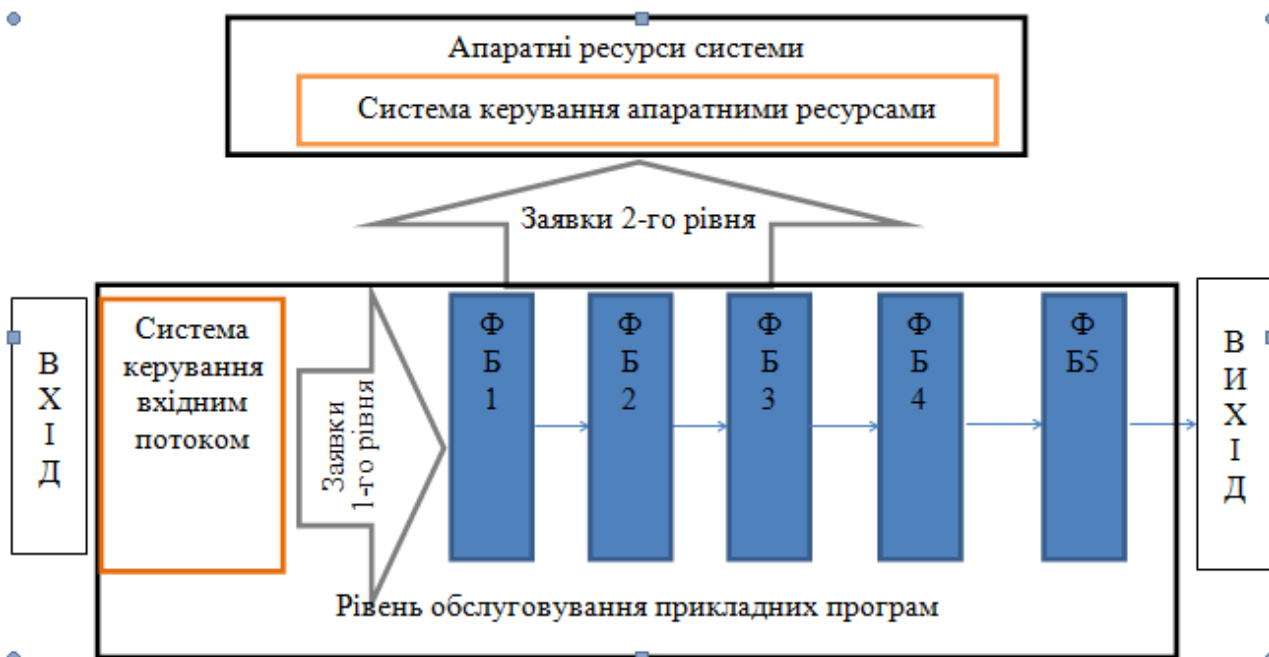


Рис. 4.14 Схема обслуговування заявок на сервері оператора мобільного зв'язку

Постає питання, яким чином організувати роботу системи керування вхідним потоком заявок, щоб потік заявок другого рівня був як найбільше рівномірним.

Обслуговування заявок першого рівня у функціональних блоках породжує потік заявок другого роду, для виконання яких використовують задану кількість ресурсів серверу. Отже, якщо у деякому функціональному блоці обробляється одночасно велика кількість заявок першого роду, при цьому заявки другого роду породжені відповідним функціональним блоком потребують для свого виконання значної кількості ресурсів, тоді може постати проблема браку ресурсів серверу, що призведе до затримки в обслуговуванні заявок першого роду, и як наслідок перевищення допустимого часу обслуговування, втрати заявок, зниження якості обслуговування абонентів.

В [53] було запропоновано стратегію керування вхідним потоком заявок, яка полягає в тому, щоб не допускати двох сплесків навантаження протягом часу обслуговуванні заявок у ресурсозатратних функціональних блоках.

Запропонований метод передбачає відслідковування піків навантаження, та введення затримки для деякої частини заявок другого піку, що дозволяє уникнути перевантаження ресурсів серверу.

З метою формування довгострокової програми згладжування навантаження було запропоновано статичний метод керування вхідним потоком заявок на тарифікацію, який передбачає розробку *схеми згладжування вхідного навантаження*.

Схема згладжування вхідного навантаження представляє собою набір значень максимально допустимої кількості заявок (послідовність $\{k_i\}$), що поступають на вхід системи за малий інтервал часу Δt_i у заданій послідовності. Кількість елементів послідовності n підбирається таким чином, щоб виконувалось рівняння

$$t = \sum_{i=1}^n \Delta t_i ,$$

де t – середній час перебування заявки першого рівня у системі.

Необхідно підібрати таку послідовність $\{k_i\}$, щоб виконувалось дві умови:

1. Заявки, які одночасно обслуговуються у системі повинні використовувати об'єм ресурсів V близький до загальної максимально можливої кількості ресурсу V_{max} . Дисперсія послідовностей таких об'ємів має бути мінімальною.

2. Дисперсія елементів послідовності $\{k_i\}$ повинна бути мінімальною.

3. Тривалість перебування заявки першого рівня у функціональних блоках (ФБ) є випадковою величиною, що залежить від швидкості обробки породжуваних у даному ФБ заявок другого рівня. Спираючись на середньо статистичні значення отримані системою моніторингу, будемо говорити, що час перебування заявки у функціональному блоці є відомим (t_j , де j - номер ФБ). $t = \sum_{j=1}^m t_j$, де m – кількість функціональних блоків у системі.

Відома кількість ресурсу (v_j , де j – номер ФБ), яка необхідна для обслуговування заявок другого рівня породжуваних заданим ФБ.

Яким чином розраховувати об'єм ресурсу V , який використовується у поточний момент часу було показано в [2], основний принцип полягає в тому, що використовується зворотня система відліку часу. Прийmemo за нульовий час закінчення обслуговування заявки, тобто $t^0 = t$, далі позначимо періоди часу коли заявки переходять між функціональними блоками: $t^1 = t^0 - t_1, \dots, t^j = t^{j-1} - t_j, \dots, t^m = t^{m-1} - t_m = 0$. Всі заявки, які надійшли протягом інтервалу $[t^1, t^0]$ в момент часу t^0 обслуговуються у першому функціональному блоці. Заявки, які надійшли до системи протягом інтервалу $[t^j, t^{j-1}]$ в момент часу t^0 , обслуговуються у j -му функціональному блоці.

Таким чином, об'єм ресурсу V^0 , що використовується у момент часу t^0 – це сума об'ємів ресурсу v_j^0 зайнятого заявками, які перебувають у j -му ФБ ($j = \overline{1, m}$) в момент часу t^0 .

$$V^0 = \sum_{j=1}^m v_j^0$$

Значення v_j^0 залежить від кількості заявок, які надійшли до системи протягом часу $[t^j, t^{j-1}]$, та визначається як добуток кількості заявок, на об'єм ресурсу, що необхідних для обслуговування однієї заявки у відповідному функціональному блоці. Оскільки застосовується *схема згладжування вхідного навантаження*, то максимально допустима кількість заявок які припадають на цей інтервал часу відома. v_j^0 – це добуток кількості заявок, що у момент часу t^0 перебувають у j -му ФБ, на об'єм ресурсу, який потрібен для обслуговування заявок другого рівня породжених j -му ФБ.

Задля забезпечення ефективного згладжування необхідно, щоб умова 1 виконувалася не тільки для об'єму V^0 , але й для всіх V^i , ($i = \overline{1, n}$).

Цільова функція має три складові:

- Дисперсія елементів послідовності $\{k_i\}$ повинна бути мінімальною.
- Дисперсія елементів послідовностей $\{V^i\}$ мінімальна

- Середнє значення елементів послідовності $\{V_j\}$ прямує до максимально можливої кількості ресурсу V_{max} заданого типу, що виділяється для обслуговування заявок вибраного типу сервісу

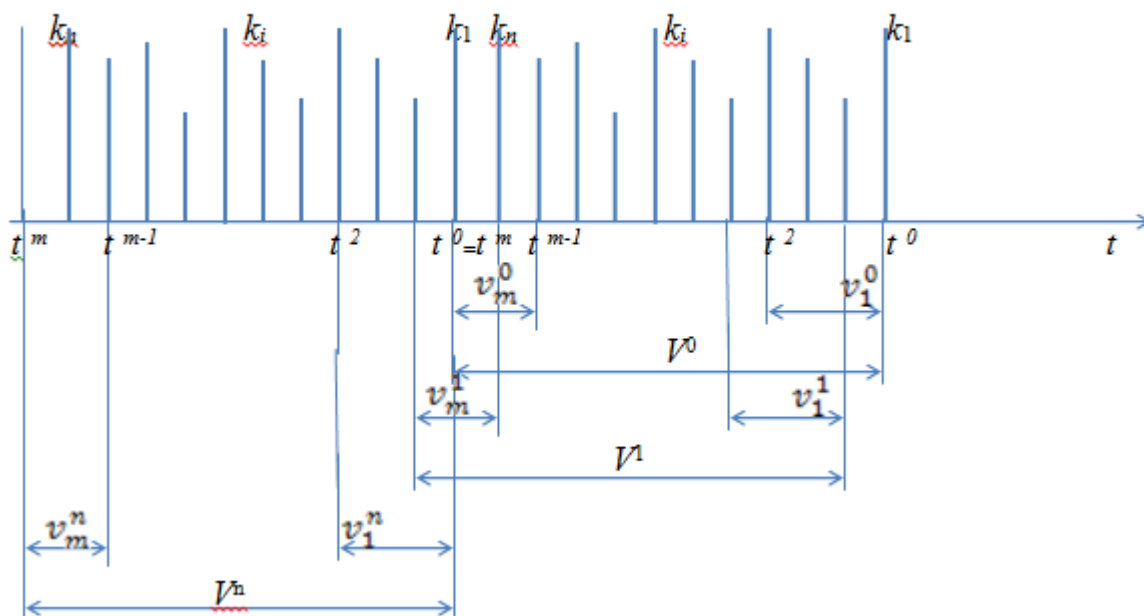


Рис. 4.15 Схема згладжування вхідного навантаження, з урахуванням об'єму ресурсу, що використовується

Методом розв'язку вибору послідовності $\{k_i\}$ здійснюється за допомогою генетичного алгоритму, таким чином щоб задовольнялися умови 1 і 2:

1. Геномом виступають елементи послідовності $\{k_i\}$.
2. Кросовер: зміна значень елементів послідовності $\{k_i\}$.
3. Умови завершення алгоритму:
 - за часом,
 - кількістю розглянутих поколінь
 - виродження популяції

В результаті отримуємо послідовність

При використанні запропонованої схеми згладжування вхідного навантаження забезпечується максимально допустима однорідність потоку

заявок другого рівня. Оскільки береться до уваги послідовність операцій, які виконуються з заявкою на сервері мобільного оператора, об'єму ресурсу, що потрібен для забезпечення цих операцій. Підбирається послідовність максимально допустимих значень кількості заявок, що надходять у систему за малий інтервал часу. Критеріями оцінки є наближеність загального об'єму, що використовується до максимально допустимого при цьому забезпечується мінімальне середньоквадратичне відхилення від середнього значення кількості заявок, що пропускаються у систему. Фактично приведено спосіб керування чергою заявок, які надходять на серверу оператора.

4.5.3. Постановка задачі методу керування вхідним потоком на сервер мобільного зв'язку.

Метод керування вхідним потоком заявок на тарифікацію, особливістю якого є контроль кількості заявок, що знаходяться на етапі обслуговування, у разі перевищення їх допустимої кількості заявки затримуються на вході, що дозволяє уникнути перевантаження ресурсів та не допустити неефективного завантаження ресурсу на обслуговування заявок, які оброблятимуться більш ніж за відведений в системі час.

Вхідними даними в задачі керування потоком заявок, які надходять на обслуговування на сервер мобільного оператора є:

- Інформація про об'єм ресурсу, який є необхідним для здійснення операцій, передбачених функціональним блоком для обслуговування заявки заданого типу сервісу.
- Інформація про тривалість використання ресурсів при обслуговуванні заявки заданого типу сервісу у кожному функціональному блоці.
- Статистична інформація про тривалість обслуговування заявки заданого типу сервісу у кожному функціональному блоці.
- Об'єм ресурсів виділених для обслуговування заданого типу сервісу.

Параметри серверу, які характеризуються як ресурси системи, що обслуговує заявки, як правило розраховані для середніх значень параметрів вхідного потоку, однак в системі наявні пікові значення кількості заявок, що надійшли одночасно.

Під сплеском навантаження вхідного потоку розуміємо одночасне надходження такої кількості заявок, яка більшою розрахованого вище допустимого значення.

Для керування процесом обробки заявок з метою запобігання дефіциту ресурсу в системі керування пропонується використання наступної стратегії:

- два і більше сплески навантаження вхідного потоку не обслуговувалися одночасно у функціональних блоках, обробка в яких потребує значної кількості ресурсів:

- для цього вводиться затримка частини заявок, надходження яких співпало зі сплеском навантаження. Час затримки визначати так, щоб затримані заявки не поступали в систему доти, доки попередній сплеск навантаження не буде успішно обслужений в ресурсозатратному функціональному блоці.

Заявки на обслуговування надходять до системи *за заданим законом*. Процес обслуговування однієї заявки включає в себе перебування (обслуговування) заявки в одному з n функціональних блоків, для обслуговування використовується G типів ресурсів. Нехай на обслуговування поступають заявки від m типів сервісів. Відома статистика часу перебування заявки i -го типу сервісу ($i = \overline{1, m}$) в j -му функціональному блоці ($j = \overline{1, n}$).

Відоме математичне очікування (t_{ij}) часу перебування заявки i -го типу сервісу ($i = \overline{1, m}$) в j -му функціональному блоці ($j = \overline{1, n}$), ці дані зведені в матрицю $T = \{ t_{ij} \}$. Відомо, що протягом обслуговування заявки i -го типу сервісу ($i = \overline{1, m}$) у j -му функціональному блоці ($j = \overline{1, n}$) ресурс g -го типу займається на час τ_{ij}^{sg} ($\tau_{ij}^{sg} \leq t_{ij}$). Інформація про тривалість обслуговування

зведена в матриці $\mathcal{T}^{Sg} = \{\tau_{ij}^{Sg}\}_{i=\overline{1,m}, j=\overline{1,n}}$. Всього таких матриць G штук, кожна матриця відповідає одному з ресурсів, що розглядається.

Відома матриця $V^{Sg} = \{v_{ij}^{Sg}\}$, кожний елемент v_{ij}^{Sg} якої відповідає об'єму ресурсу g -го типу який використовується при обслуговуванні заявки i -го типу ($i = \overline{1,m}$) в j -му функціональному блоці ($j = \overline{1,n}$).

В рамках цього дослідження не розглядається деталізація бізнес процесів, які відбуваються у функціональному блоці, тобто не уточнюється на якому саме етапі обслуговування заявки в середині функціонального блока який ресурс використовується. Тому зроблено припущення: всі заявки які у поточний час обслуговуються у функціональному блоці використовують ресурси рівномірно, тобто об'єм g -го ресурсу, що використовується i -м типом сервісу у j -му ФБ зменшується пропорційно відношенню часу використання ресурсу до часу перебування заявки у функціональному блоці, тоді справедлива формула: $v_{ij\ new}^{Sg} = v_{ij}^{Sg} \frac{\tau_{ij}^{Sg}}{t_{ij}}$, де $v_{ij\ new}^{Sg}$ – індексований об'єм ресурсу g -го типу, який використовується протягом часу обслуговування заявки i -го типу сервісу у j -му ФБ. Формуються нові матриці $V_{new}^{Sg} = \{v_{ij\ new}^{Sg}\}$.

Необхідно визначити метод керування вхідним потоком заявок, що дозволяє уникнути дефіциту ресурсів системи.

Алгоритм методу. Розшифрування позначень застосованих у алгоритмі наводиться після нього.

1. Для кожного i -го типу сервісу задати допустиму кількість заявок які можуть одночасно поступати на обслуговування в систему ($k_{i\ доп}$). Кількість допустимих заявок залежить від інтервалу дискретизації часу, система відліку дискретного часу має бути єдиною для всієї системи. *Зауваження.* У подальших роботах буде розглянутий метод визначення допустимої кількості

заявок, що розв'язується як задача динамічного програмування (задача про загрузку машини).

2. Задається множина $F=\{\emptyset\}$. Для кожного типу ресурсу $g = \overline{1, G}$ в матриці V_{new}^{Sg} знаходиться максимальний елемент $v_{i_g j_g}^{Sg} = \max\{v_{11 new}^{Sg}, v_{12 new}^{Sg}, \dots, v_{mn new}^{Sg}\}$, пари індексів $(i_{g1} j_{g1})$ відповідних елементів додаються до множини F . Якщо у матриці присутні два або більше ($gmax \geq 1$) максимальних елементи $v_{i_{g1} j_{g1}}^{Sg} = \dots = v_{i_{g gmax} j_{g gmax}}^{Sg}$, тоді в множину F додаються всі пари індексів, та позначаються $(i_{g1} j_{g1}, \dots, i_{g gmax} j_{g gmax})$. Індеси максимальних значень об'єму для різних типів ресурсів можуть співпадати; значення, що повторюються, до множини F не додаються. Наприклад, $i_{11} j_{11} = i_{21} j_{21} = 2\ 3$, це означає, що для ресурсу 1 і для ресурсу 2 перший максимальний елемент відповідає процесу обслуговування заявки 2-го типу сервісу у третьому функціональному блоці, тобто це обслуговування є найбільш витратним для ресурсів першого та другого типу, в такому випадку пара (2,3) увійде до множини F один раз. Таким чином, множина F заповнюється парами, де на першій позиції стоїть номер сервісу, обслуговування якого є ресурсозатратним у функціональному блоці, номер якого стоїть на другій позиції. *Зауваження.* Пари номерів не зберігають тип ресурсу, оскільки це не має значення для даного методу керування.

3. Елементи множини F впорядковуються за першим елементом. Множина F розділяється на m підмножин, таким чином, щоб в F_1 увійшли пари де перший елемент дорівнює 1, в F_2 увійшли пари де перший елемент дорівнює 2, тощо. Якщо деяка r -та підмножина ($r \in \overline{1, m}$) буде порожньою ($F_r = \{\emptyset\}$), тоді для заявок r -го типу сервісу не будуть застосовуватися затримки заявок, що надійшли у сплесках навантаження вхідного потоку. Для всіх підмножин F_d ($d \in \overline{1, m}$), де міститься один елемент, виконуються дії п.4. Для всіх підмножин F_p ($p \in \overline{1, m}$), де міститься два і більше елементи, виконуються дії з п.5.

4. Завдання цього пункту полягає в тому, щоб визначити максимальну затримку надлишкової кількості заявок d -го типу сервісу, які надійшли у моменти пікових навантажень вхідного потоку. Якщо в множині F_d міститься 1 елемент (d, f_d) , це означає, що для заявки d -го типу сервісу не можна допускати двох піків навантаження протягом часу обслуговування у функціональному блоці f_d , тривалість якого визначається з матриці T та дорівнює $t_{d f_d}$. Перехід до п. 6.

5. Завдання цього пункту не тільки не допустити припадання двох піків навантаження на один критичний (ресурсозатратний) функціональний блок, але й уникнути суперпозиції, коли два піки навантаження обслуговуються у двох ресурсозатратних функціональних блоках. Для цього елементи підмножини F_p впорядковуються за другим елементом. Припустимо, що множина F_p складається з двох елементів: (p, f_{1p}) і (p, f_{2p}) , задачі з більшою кількістю елементів мало ймовірні та вирішуються у аналогічний спосіб. Це означає, що при обслуговуванні заявок p -го типу сервісу, ресурсозатратними є функціональні блоки з номерами f_{1p} і f_{2p} . З матриці T обираються елементи з відповідними індексами: $t_{p f_{1p}}$, $t_{p f_{2p}}$. Умови при яких два піки навантаження не припадуть на один функціональний блок наступні:

А) відстань між піками навантаження не може бути меншою ніж значення $t_{p f_{1p}}$.

В) відстань між піками навантаження не може бути меншою ніж значення $t_{p f_{2p}}$.

С) Якщо $f_{2p} - f_{1p} = x > 1$, тоді не допускається відстань між піками навантаження більша ніж $\sum_{q=1}^{x-1} t_{p (f_{1p}+q)}$.

6. В процесі роботи системи моніторингу фіксуються моменти пікових навантажень, коли у систему надійшла кількість заявок, що є більшою за допустиме значення (відповідно до п. 1). Моменти часу, коли виявлено сплеск навантаження, додаються до множин $T_{i \max}$, де i – тип сервісу, для

якого зафіксовано сплеск навантаження. Для сервісів r -того типу (див. п.3) множина $T_{r \max}$ не створюється.

Для сервісів типу d , для елементів множини $T_{d \max}$ перевіряється умова п.4, тобто для кожного нового елементу множини $t_{d \max w+1}$ перевіряється значення $t_{d f_d} - (t_{d \max w+1} - t_{d \max w}) = y1$, якщо $y1 > 0$, тоді частина заявок $(k_d(t_{d \max w+1}) - k_{d \text{ доп}})$ затримується на час $y1$. Якщо у момент часу $(t_{d \max w+1} + y1)$, кількість заявок що надійшла $k(t_{d \max w+1} + y1)$ плюс залишок $(k_d(t_{d \max w+1}) - k_{d \text{ доп}})$ в сумі дають значення більше допустимого $k_{d \text{ доп}}$. Тоді надлишок передається у наступний момент часу $(t_{d \max w+1} + y1 + 1)$, процедура згладжування навантаження.

Для сервісів типу p , для елементів множини $T_{p \max}$ перевіряються умови п.5, тобто для кожного нового елементу множини $t_{p \max w+1}$ перевіряються умови:

- значення $t_{p f_{1p}} - (t_{p \max w+1} - t_{p \max w}) = y2$, якщо $y2 > 0$, тоді частина заявок $(k_p(t_{p \max w+1}) - k_{p \text{ доп}})$ затримується на час $y2$, у разі потреби застосовується процедура згладжування навантаження

- значення $t_{p f_{2p}} - (t_{p \max w+1} - t_{p \max w}) = y3$, якщо $y3 > 0$, тоді частина заявок $(k_p(t_{p \max w+1}) - k_{p \text{ доп}})$ затримується на час $y2$, у разі потреби застосовується процедура згладжування навантаження.

- якщо $f_{2p} - f_{1p} = x > 1$, тоді досліджується значення $\sum_{q=1}^{x-1} t_{p(f_{1p}+q)} - (t_{p \max w+1} - t_{p \max w}) = y4$, якщо $y4 < 0$, тоді частина заявок $(k_p(t_{p \max w+1}) - k_{p \text{ доп}})$ затримується на час $(t_{p f_{2p}} + y4)$, у разі потреби застосовується процедура згладжування навантаження.

Таким чином, у разі реєстрації події надходження другого піку навантаження протягом часу, який визначено умовами, здійснюється затримка надлишкової кількості заявок на час визначений алгоритмом методу, після чого затримані заявки надсилаються в систему так, щоб не допустити створення піку навантаження.

Алгоритм методу керування наведено на Рис. 4.16.

Кількість заявок, що є допустимою, для заданого типу сервісу розраховується методом перерозподілу технічних засобів між заявками різних типів послуг описаним вище, при цьому враховується ефективність обслуговування всіх типів сервісів при наявному об'ємі ресурсів системи.

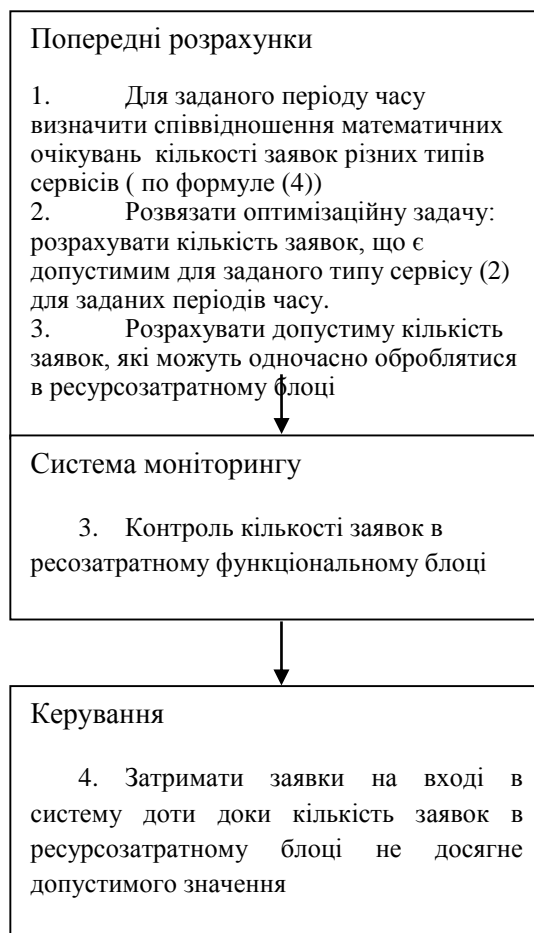


Рис. 4.16 Алгоритм методу керування вхідним потоком заявок для запобігання дефіциту серверних ресурсів

4.5.4. Моделювання методу динамічного керування вхідним навантаженням.

Проведено імітаційне моделювання методу керування потоком заявок на тарифікацію. Для моделювання було використано пакет GPSS.

В процесі імітаційного моделювання досліджувалася модель для двох ресурсів і потоку сервісів. Ресурси, що враховувалися під час моделювання – RAM and Permanent storage.

Процес обробки заявки включає в себе чотири функціональні блоки. Робота функціональних блоків імітувала такі операції як: вилучення інформації абонента з бази даних, розрахунок вартості послуги, формування нотифікації для абонента, фінальний підрахунок та списання коштів.

Для забезпечення обслуговування був виділений заданий об'єм ресурсів, розрахований на одночасне обслуговування 50 тисяч заявок на тарифікацію, за умови рівномірного розподілу кількості заявок між функціональними блоками. Під час обслуговування заявки в функціональному блоці відповідна кількість ресурсу блокувалася, та звільнялася при переході до наступного функціонального блоку. Якщо заявка надходить на обслуговування у функціональний блок, але ресурсу не достатньо для здійснення обслуговування, то заявка затримується до звільнення необхідної кількості ресурсу. На кожному етапі перевіряється час перебування заявки в системі та порівнюється з допустимим часом обслуговування. Значення були обрані максимально наближеними до реальних систем.

Вхідний потік був змодельований за законом Пуасона. Виходячи з аналізу роботи реальних системи найбільше ресурсів витрачається під час формування повідомлення абоненту. Тому в моделі здійснювався контроль за кількістю заявок, які обслуговувалися у поточний момент часу в третьому функціональному блоці та здійснювалася затримка повідомлень доти, доки кількість заявок не стане меншою за максимально допустиму кількість.

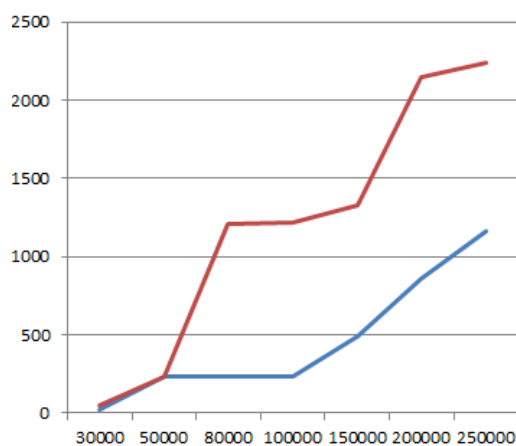


Рис. 4.17 Кількість втрачених заявок

Динаміку залежності кількості втрачених заявок від інтенсивності вхідного потоку показано на Рис. 4.17. З Рис. 4.17 видно зменшення втрат заявок через перевищення допустимого часу обслуговування. Червоним позначена лінія втрат пакетів без застосування запропонованого методу керування вхідним потоком, синім позначені результати моделювання за допомогою запропонованого методу керування.

4.6. Задача реконфігурації гетерогенного телекомунікаційного середовища

4.6.1. Метод відображення віртуальних вузлів на фізичні вузли

В основі запропонованого методу лежить спільне розташування індивідуальних ланцюгів сервісів ядра мережі на фізичній мережі. Під ланцюгом сервісів ядра мережі розуміється послідовність мережевих функцій мобільного ядра, яку повинен пройти потік трафіку [104]. Припустимо, що функціональність віртуальних мережевих функцій мобільного ядра і їх інтерфейси аналогічні функціям і інтерфейсам мережевих елементів ядра архітектури 3GPP LTE Evolved Packet Core.

Фізична мережа має вигляд графа $SN = (N, L)$ з множиною фізичних вузлів (N) і множиною каналів (L). Кожен канал $l = (n_1, n_2) \in L$, $n_1, n_2 \in N$ має максимальну пропускну здатність $c(n_1, n_2)$, а кожен вузол $n \in N$ пов'язаний з певними ресурсами c_n^i , $i \in R$, де R – множина усіх типів ресурсів. Множина точок агрегації трафіку (Traffic Aggregation Point – TAP) мережі позначається $T \subseteq N$. Для кожного вузла множини $virt_n$ – бінарний параметр, який визначає, чи є той чи інший вузол віртуальним; $phys_n^j$ – бінарний параметр, який визначає, чи є вузол n виділеним апаратним блоком функції типу $j \in F$ (F – множина типів мережевих функцій). Віртуальне ядро мережі являє собою множину ланцюгів сервісів, вбудованих в фізичну мережу.

Для каналу між двома функціями (j_1 і j_2) вимоги смуги пропускання, що відносяться до ланцюга сервісів ТАР $t \in T$ позначаються як $d_t^{(j_1, j_2)}$. $d_t^{j, i}$ – кількість ресурсу типу i , що виділяється для мережевої функції j ТАР t . $s_t^{j, i}$ позначає час обробки запиту на ресурсі типу i мережевої функції j ТАР t однією одиницею ресурсу для випадку віртуальної мережевої функції. Задано значення оброблюваних запитів в секунду виділеного апаратного функціонального блоку n типу j , і позначено як μc_n^j . Вимоги до запитів в секунду мережевої функції j , що відноситься до ланцюга сервісів ТАР t , позначаються як M_t^j .

Оптимізація мережі передбачає знаходження розташування мережевих функцій ядра і розподіл ресурсів, які дозволяють мінімізувати витрати на зайняті ресурси вузлів і каналів фізичної мережі, а також вимоги трафіку. Цільова функція буде мати вигляд лінійної комбінації з ваговими коефіцієнтами a , b і c (4.3) трьох вартісних виразів: базова вартість $cost(n)$ – використовується у випадку, якщо деяка мережева функція розміщується на фізичному вузлі $n \in N$, вартість зайнятої одиниці ресурсів $cost(i, n)$ на фізичному вузлі n і вартість зайнятої одиниці пропускну здатності $cost(n_1, n_2)$ на фізичному каналі $(n_1, n_2) \in L$.

Формули (4.3)-(4.12) являють собою постановку оптимізаційної задачі нелінійного програмування. Змінні $x_n^{t, j}$ використовують для визначення того, чи розташована мережева функція j на фізичному вузлі n . Для $j = \text{ТАР}$, $x_n^{t, \text{ТАР}}$ – не змінні, а вхідні параметри, які вказують де знаходиться ТАР $t \in T$:

$$x_n^{t, \text{ТАР}} = \begin{cases} 1 & \text{if } t = n, \\ 0 & \text{else.} \end{cases}$$

Змінні $f_{(n_1, n_2)}^{t, (j_1, j_2)}$ вказують, чи фізичний канал $(n_1, n_2) \in L$ використовується для шляху між j_1 і j_2 для ТАР $t \in T$.

$$\min_{x_n^{t,j}, f_{(n_1, n_2)}^{t, (j_1, j_2)}, d_t^{j, i}} (a \cdot \sum_{n \in N} x_n \cdot cost(n) + b \cdot \sum_{n \in N} \sum_{t \in T} \sum_{j \in V} \sum_{i \in R} x_n^{t, j} \cdot d_t^{j, i} \cdot cost(i, n) + c \cdot$$
(4.3)

$$\sum_{(n_1, n_2) \in L} cost(n_1, n_2) \cdot \sum_{t \in T} \sum_{(j_1, j_2) \in E} f_{(n_1, n_2)}^{t, (j_1, j_2)} \cdot d_t^{(j_1, j_2)})$$

Subject to $\sum_{n \in N} x_n^{t, j} = 1 \quad \forall t \in T, j \in V$

(4.4)

$$x_n^{t, j} \leq d_t^{j, i} \quad \forall t \in T, j \in V, n \in N, i \in \{R \setminus bdw\}$$
(4.5)

$$\sum_{(w, n) \in L} \sum_{t \in T} \sum_{(j_1, j_2) \in E} f_{(w, n)}^{t, (j_1, j_2)} \cdot d_t^{(j_1, j_2)} \leq c_n^{bdw} \quad \forall n \in N$$
(4.6)

$$\sum_{t \in T} \sum_{j \in V} x_n^{t, j} \cdot d_t^{j, i} \leq c_n^i \quad \forall n \in N, i \in \{R \setminus bdw\}$$
(4.7)

$$\sum_{t \in T} \sum_{(j_1, j_2) \in E} f_{(n_1, n_2)}^{t, (j_1, j_2)} \cdot d_t^{(j_1, j_2)} \leq c(n_1, n_2) \quad \forall (n_1, n_2) \in L$$
(4.8)

$$\sum_{(n, w) \in L} f_{(w, n)}^{t, (j_1, j_2)} - f_{(n, w)}^{t, (j_1, j_2)} = x_n^{t, j_1} - x_n^{t, j_2} \quad \forall t \in T, n \in N, (j_1, j_2) \in E$$
(4.9)

$$x_n^{t, j}, f_{(n_1, n_2)}^{t, (j_1, j_2)} \in \{0, 1\} \quad \forall t \in T, j \in V, n \in N, (j_1, j_2) \in E, (n_1, n_2) \in L$$
(4.10)

$$\mu_t^j \geq M_t^j \quad \forall t \in T, j \in V$$
(4.11)

$$\mu_t^j = \sum_{n \in N} \left(x_n^{t, j} \cdot \sum_{i \in R} \frac{d_t^{j, i}}{s_t^{j, i}} \cdot virt_n + x_n^{t, j} \cdot \mu c_n^j \cdot phys_n^j \right) \quad \forall t \in T, j \in V$$
(4.12)

$$\sum_{(j_1, j_2) \in E} \sum_{(n_1, n_2) \in L} f_{(n_1, n_2)}^{t, (j_1, j_2)} \cdot L(n_1, n_2) \leq L_t \quad \forall t \in T$$

Булеві змінні x_n позначають, чи розміщується мережева функція на вузлі $n \in N$. Якщо $x_n=0$, жодна функція із жодного ланцюга сервісів не розміщується на даному вузлі. Формула (4.4) означає, що для кожної ТАР/ланцюга сервісів розміщується лише одна функція кожного типу. Вираз (4.5) гарантує, що усі ресурси розташовуються на фізичних вузлах, обраних для розташування відповідних мережевих функцій.

Формули (4.6), (4.7) і (4.8) накладають обмеження на ресурси каналів і вузлів мережі. Канал, який сполучає дві мережеві функції, являє собою шлях

фізичної мережі. Його вимоги до смуги пропускання впливають як на мережеві ресурси фізичних вузлів, на яких розміщені функції, так і на мережеві ресурси проміжних вузлів, що лежать на їхньому шляху. Формула (4.9) являє собою обмеження щодо збереження потоку для всіх шляхів у фізичній мережі. Формула (4.10) гарантує, що всі змінні задачі відображення шляху розміщення мережевої функції є булевими.

Для врахування продуктивності моделі, у виразі (4.11) описано обмеження на кількість запитів за секунду. Ці обмеження гарантують, що загальна швидкість обслуговування мережевої функції (μ_t^j) перевищує деяку порогову величину. Для обмеження затримок на каналах, додається відповідне значення (4.12), яке обмежує загальну затримку на шляху. На Рис. 4.18 зображено приклад системи розподілу запитів мережі.

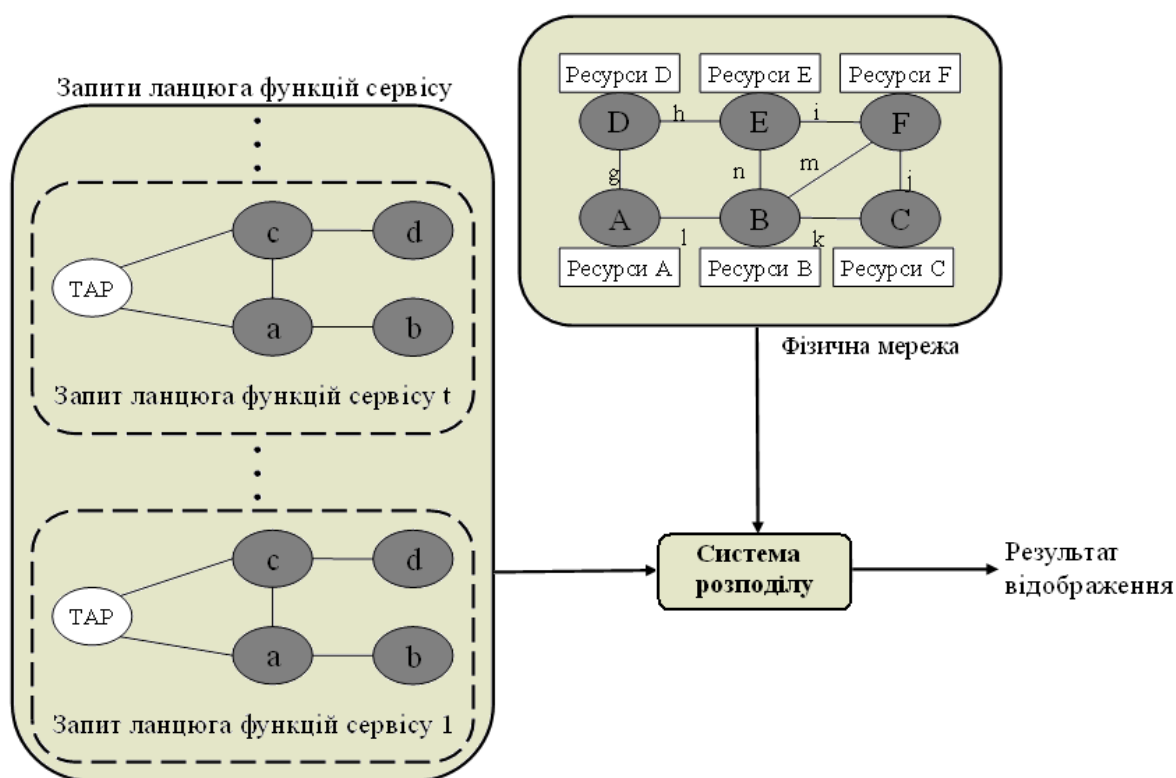


Рис. 4.18 Приклад топології системи виділення ресурсів мережі

Задачу (4.3) – (4.12) планується вирішувати в режимі офлайн. Відповідно до прийнятого рішення, для кожного з блоків мережевої функції виділяється деяка кількість ресурсів, на основі грубої оцінки його потреби.

Миттєві потреби задовольняються динамічно, піз час кожного виконання мережевої функції.

Для оцінки запропонованого алгоритму, було здійснено моделювання генетичного алгоритму в системі MATLAB. На Рис. 4.19 представлено фрагмент лістингу програми.

```
ObjectiveFunction=@fitness;
>>nvars=108;% Number of variables
>>LB=zeros(1,108);% Lower bound
>>UB=ones(1,90);% Upper bound
>>IntCon=ones(1,90);% Integer variables
>>ConstraintFunction=@constraint;
>>[x,fval]=ga(ObjectiveFunction,nvars,[],[],[],LB,UB,ConstraintFunction,IntCon,
options)
```

Рис. 4.19 Фрагмент коду MATLAB для оцінки запропонованого алгоритму

Для моделювання було використано приклад простої системи з 10-ма вузлами, 3-ма функціональними блоками і 2-ма типами ресурсів. Результати у вигляді кривої, на якій зображено Зміну загальних витрат при зміні вимог до обслуговування вхідного навантаження, зображено на Рис. 4.20.



Рис. 4.20 Кількість витрат на розміщення вузлів при динамічній зміні вимог до інтенсивності обслуговування.

З графіка видно, що вартість оптимального виділення ресурсів відчутно змінюється при зміні потрібної інтенсивності обслуговування запитів. Таким чином, вона залежить від інтенсивності надходження запитів на обслуговування протягом дня.

4.6.2. Алгоритм динамічного розподілу мережесих ресурсів

Припустимо, що рішення щодо розміщення функціональних блоків мережі знайдено за допомогою виразів (4.3) – (4.12). Для прикладу розглянемо випадок, коли кількість ресурсів, виділених для блоку мережевої функції, потрібно збільшити. Зміна у вимогах до обслуговування є заданою величиною.

Мета застосування алгоритму розподілу ресурсів (Рис. 4.21) полягає у визначенні кількості та місця розташування ресурсів, які повинні бути додані до мережевої функції. Запропонований алгоритм складається з трьох основних кроків.

Крок 1. Перевірка можливості виділення потрібної кількості ресурсів на вузлі поточного розташування функціонального блоку j для ТАР t .

Крок 2. Якщо кількість вузлів, визначених у першому кроці є недостатньою, здійснюється перевірка наявності необхідної кількості ресурсів на вузлах розгортання функціональних блоків j для інших ТАР.

Крок 3. Якщо кількість вузлів, визначених у другому кроці є недостатньою, перевіряється решта вузлів.

Якщо на вузлі, відмінному від поточного (кроки 2 і 3), знайдено необхідну кількість ресурсів для функціонального блоку, здійснюється міграція функції на цей вузол. Якщо необхідно відняти ресурси, виконується просте звільнення ресурсів. Алгоритм динамічного виділення ресурсів зображено на Рис. 4.21.

На рис. 4.21:

$\Delta d_t^{j,i}$ – зміна вимог до ресурсів;

$d'_t{}^{j,i} = d_t{}^{j,i} + \Delta d_t{}^{j,i}$, $\Delta c_n{}^i$ – залишкові ресурси типу i на вузлі n ;

$\Delta \mu_t{}^j$ – різниця вимог до обслуговування між попереднім і наступним моментом часу функціонального блоку j для ТАР t ;

$y^{t,j}$ – розташування функціонального блоку j для ТАР t .

for all t
 for all $j: \mu_t^j \leq M_t^j$
 if $\exists \Delta d_t^{j,i}: \Delta d_t^{j,i} \leq \Delta c_n^i \forall i \in R, \Delta \mu_t^j \geq \sum_{i \in R} \frac{\Delta d_t^{j,i}}{s_t^{j,i}},$ where
 $n: x_n^{t,j} = 1$
 $y^{t,j} = n$
 end if
 if
 for all $n2: \exists t2: x_{n2}^{t2,j} = 1$
 $\exists d_t^{j,i}: d_t^{j,i} \leq \Delta c_{n2}^i \forall i \in R, \Delta \mu_t^j \geq \sum_{i \in R} \frac{\Delta d_t^{j,i}}{s_t^{j,i}}$
 $y^{t,j} = n2$
 end for
 end if
 if
 for all $n3: \nexists t3: x_{n3}^{t3,j} = 1$
 $\exists d_t^{j,i}: d_t^{j,i} \leq \Delta c_{n3}^i \forall i \in R, \Delta \mu_t^j \geq \sum_{i \in R} \frac{\Delta d_t^{j,i}}{s_t^{j,i}}$
 $y^{t,j} = n3$
 end for
 end if
 end for
 end for

Рис. 4.21 Алгоритм динамічного виділення ресурсів

Загальна схема процесу виділення ресурсів зображена на Рис. 4.22.

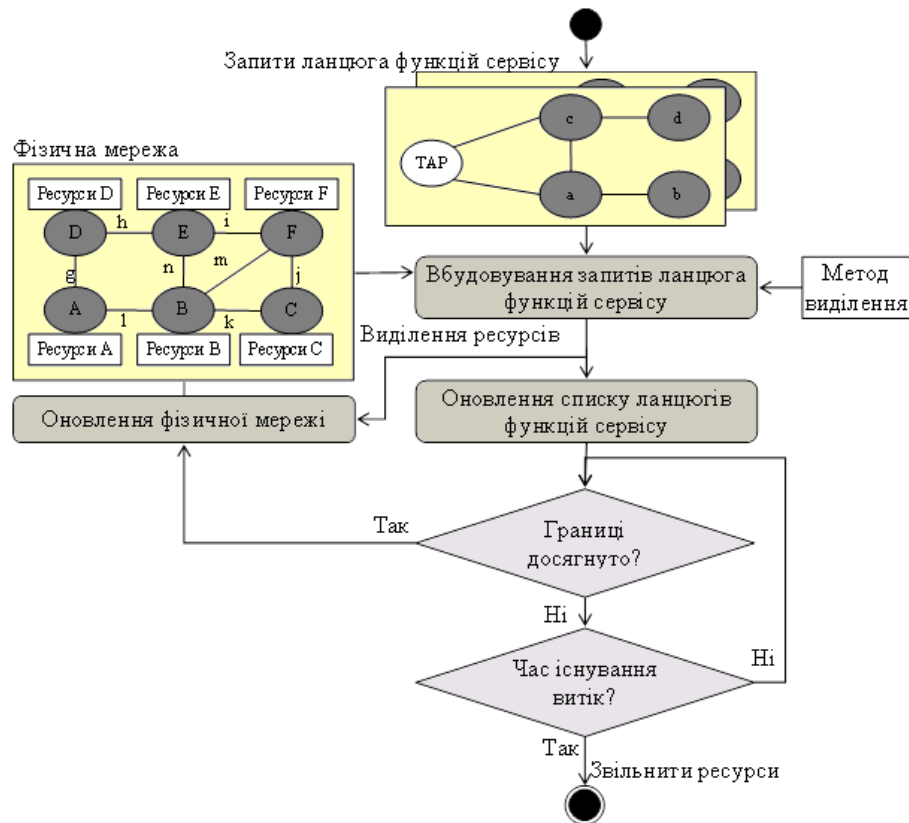


Рис. 4.22 Діаграма активності: життєвий цикл запиту мережі

Початок життєвого циклу мережевого запиту починається з приходом запиту ланцюга функцій сервісу (Рис. 4.22). На початковому етапі метод виділення ресурсів застосовується для вбудовування запитів функцій мережі. В якості вхідних даних він приймає поточний стан мережі (доступні ресурси пам'яті, CPU та пропускної здатності), а також самі запити. Коли надходить новий запит, і досягається межа рівня обслуговування, використовується метод виділення ресурсів з Рис. 4.21 для вбудовування запитів.

Метою запропонованого методу виділення ресурсів є мінімізація їх використання при забезпеченні заданого рівня якості обслуговування. Він надає телекомунікаційному оператору можливість мінімізувати операційні та капітальні витрати за рахунок використання парадигми хмарних обчислень, а також покращити якість сприйняття.

4.6.3. Принцип роботи методу відображення і планування мережевих функцій

Задача відображення і планування мережевих функцій (Network Function Mapping And Scheduling – NFMS) має багато можливостей для спільного використання мережевих ресурсів. Одна з них полягає у тому, що для кожного VNF використовується виділена віртуальна машина. В такому випадку фізичні ресурси будуть швидко закінчуватись, витрачаючись на «легкі» функції, які могли б бути оброблені однією віртуальною машиною (наприклад, за допомогою контейнерів). Розглянемо підхід до спільного використання ресурсів, завдяки якому задана віртуальна машина може обробляти декілька VNF з черги (наприклад, один за одним).

NFMS складається з необхідності онлайн обробки мережевих сервісів з використанням набору із f віртуальних мережевих вузлів: $F = \{1, ..., f\}$. Це означає, що кожен сервіс створюється і вбудовується по мірі його необхідності. Усі сервіси мережі S складаються з послідовності $SF = \{1, ..., m\}$ з m VNF, де функція $1 \leq i \leq m$ повинна оброблятися на множині $F(i) \subseteq F$

вузлів. Функції $\{1, \dots, m\}$ обробляються в певній послідовності, при цьому один віртуальний вузол може обробляти не більше однієї функції за раз. Час обробки для функції i на вузлі $j \in F(i)$ повинен бути $\rho_{ij} > 0$, де $1 \leq j \leq f$, і під час обробки або в черзі для обробки, функція i використовує буфер δ_i на вузлі, на який вона відображається. В будь-який момент часу розмір буфера для заданого вузла j дорівнює B_j .

Визначимо бінарну змінну $\beta_{i,j}$ для кожної комбінації функцій і вузлів.

$$\beta_{i,j} = \begin{cases} 1, & \text{якщо вузол } j \text{ може обробляти функцію } i \\ 0, & \text{в іншому випадку} \end{cases}$$

Основна задача полягає у тому, щоб вибрати для кожної VNF i віртуального вузла $j \in F(i)$ і часу завершення t_i , коли її обробка буде завершена, тобто час t_s , при якому обробка функції i на вузлі j починається, може бути отриманий з $t_s = t_i - \rho_{ij}$. Також знайдемо граничний термін обробки функції (t_l). Після цього для кожного віртуального вузла j , визначимо очікуваний час завершення π_j останньої функції у черзі на обробку на вузлі, і для кожного сервісу визначаємо час, коли запит на відображення і планування сервісу отримується фізичною мережею – час прибуття t_a .

Таким чином, задача NFMS складається з двох частин:

1. Проблема відображення – визначення віртуальних вузлів, на які повинна бути відображені кожна VNF.
2. Проблема планування – визначення порядку обробки відображених VNF для кожного вузла.

На Рис. 4.23 та Рис. 4.24 представлено алгоритм відображення та планування віртуальних функцій мережі. На Рис. 4.23 показується псевдокод алгоритму, де S позначає критерій, за яким ранжуються вузли.

```

Старт
Здійснити резервне копіювання стану фізичної мережі
for  $i \in S$  do
    Ініціалізація: Множина можливих вузлів  $F' = \emptyset$ 
    if  $(i=1)$  then
         $t_{i-1} = t_a$ 
    end if

```

```

for  $j \in F$  do
 $t_e = \rho_{ij} + \max(\pi_j, t_{i-1})$ 
if  $((\beta_{ij} = 1) \wedge (B_j \geq \delta_i) \wedge (t_e \leq t_i))$  then
 $F' = F' \cup f$ 
end if
end for
if  $F' \equiv 0$  then
Неуспіх відображення та планування
Додавання VNF
return
end if
Відсортувати  $F'$  відповідно до  $C$ 
Обрати найвищий вузол  $j^*$  з  $F'$ 
Відобразити функцію  $i$  на  $j^*$ 
Встановити  $t_i = \max(\pi_j, t_{i-1})$ 
Оновити  $B_j, \pi_j, t_{i-1}$ 
end for
Відображення та планування завершено
Кінець

```

Рис. 4.23 Алгоритм відображення мережових функцій (жадібне відображення функцій)

$$F = F \cup (f+1)$$

$$B_{f+1} = \delta_i$$

$$\pi_{f+1} = 0$$

Рис. 4.24 Алгоритм планування мережових функцій (додавання VNF)
Алгоритм відображення та планування VNF

1. Для кожної функції i визначаються всі вузли $F(i) \subseteq F$, які беруть участь у її обробці.
2. Вузли $F(i)$ ранжуються на основі жадібного критерію.
3. Для відображення обирається вузол з найкращим рангом; функція планується на обробку в кінці черги вузла.
4. Визначається фактичний час початку обробки на основі доступності вузла та завершення обробки попередньої функції.

На кожному етапі планування потрібно визначати завершеність обробки, щоб переконатись, що воно знаходиться в межах встановленого граничного терміну для сервісу.

Якщо під час відображення останньої функції сервісу час завершення перевищує граничний термін, або функція не має жодного вузла-кандидата (тобто вони всі завантажені), то додавання ресурсів пропонується здійснювати відповідно до алгоритму на Рис. 4.24.

Запропонований алгоритм базується на тому, що функції відображаються на вузли таким чином, щоб їх завантаженість була близькою до оптимальної за критерієм енергоефективності значення.

4.7. Метод “нескінченний потяг” організації обчислень у хмарному середовищі для гібридних телекомунікаційних сервісів

На сьогоднішній день існує декілька підходів до організації обчислювального процесу в контролері SDN. Їх основним недоліком є те, що вони не враховують можливість застосування необмеженого ресурсу, який надають хмарні технології. Усі існуючі підходи засновані на тому, що кількість ресурсів системи є обмеженою і залежить від технічних можливостей сервера, на якому розміщуються віртуальні машини.

Наслідком застосування даних підходів стає зниження якості надання послуг в певні періоди часу, які пов’язані з тим, що робота системи моніторингу полягає у збиранні та аналізі даних про якість надання послуг в мережі. Якщо значення параметрів QoS для певних послуг є нижче порогового, але інші послуги надаються з досить високою якістю, система може не реагувати на зниження якості у певні періоди часу, так як враховується лише середнє значення показників.

Вирішенням даної проблеми є застосування гетерогенного хмарного середовища, а саме, певних рішень для платформи як інфраструктури (PaaS). Platform-as-Infrastructure являє собою ізольований кластер, що складається з групи серверів і сервісів, які взаємодіють як цілісна система, надаючи

можливість зручно розгортати, тестувати, підтримувати і масштабувати систему. PaaS дозволяє створювати і обслуговувати необмежену кількість віртуальних машин, робота яких полягає в обслуговуванні обчислювальних функцій контролера. Застосування необмеженої кількості ресурсів для роботи віртуальних машин дозволить уникнути періодів зниження якості надання послуг.

Метод, що пропонується в даній статті, називається «ефект нескінченного потягу». Його основна ідея полягає в тому, що після того, як певна віртуальна машина отримала на обслуговування задану кількість задач, створюється нова віртуальна машина, на яку надходять усі наступні задачі (Рис. 4.25).

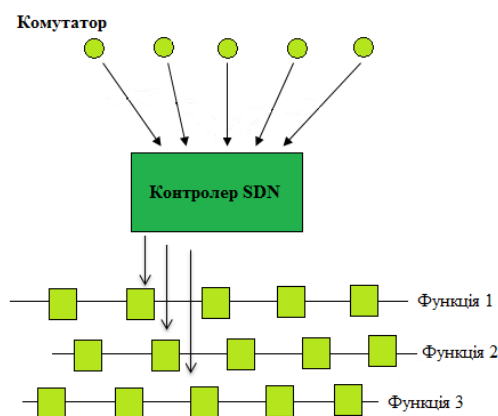


Рис. 4.25 Метод «нескінченного потягу»

Запропонований метод спирається на метод динамічної міграції віртуальних машин, розроблений компанією Jelastic [11], що забезпечує балансування навантаження серверів хмарного сховища даних за рахунок створення платформи для автоматичного керування контейнерами віртуальних машин. Також враховується метод вибору контейнера для міграції віртуальних машин, описаний в статті [12], який дозволяє аналізувати та прогнозувати навантаження на мережу на основі оцінки використовуваних ресурсів. Застосування хмарного сховища даних передбачає, що усі дані зберігаються на великій кількості розподілених в мережі серверів. При цьому повинні виконуватись дві основних задачі:

- Інтерактивний розподіл завдань клієнтів між віртуальними машинами, які розташовані в одному або декількох кластерах. Це з одного боку завдання балансування навантаження, а з іншого – завдання забезпечення надійності обслуговування клієнтів;

- Контроль і керування роботою кластера віртуальних машин. Ресурсів кластера має завжди вистачати всім віртуальним машинам, які одночасно працюють на всіх серверах кластера.

Метод «нескінченного потягу» передбачає, що всі заявки на обслуговування направляються до поточної віртуальної машини до тих пір, поки вона не буде заповнена. Кількість заявок, які можуть бути оброблені однією віртуальною машиною залежить від кількості ресурсів, яка виділяється їй при створенні. Після заповнення даної машини створюється нова, на яку перенаправляються усі наступні заявки. Використання даного методу є можливим в тому випадку, якщо кількість ресурсів системи є умовно необмеженою. Використання гетерогенного хмарного середовища надає таку можливість.

Старі віртуальні машини продовжують обслуговувати свої інформаційні потоки до тих пір, поки їх кількість не вичерпається. Після цього віртуальні машини згортаються, або очікують на повторне введення в експлуатацію у якості «пустих вагонів». Максимальна кількість заявок, які можуть обслуговуватись у «вагоні», залежить від конфігурації хмарної платформи, або може бути отримана дослідним шляхом. Кількість заявок визначається об'ємом ресурсу, який використовує контролер для виконання обчислювальних завдань. На максимальну кількість заявок також впливає гнучкість процесів міграції – технічних процесів обслуговування віртуальних машин. Ілюстрація описаного методу зображена на Рис. 4.26.

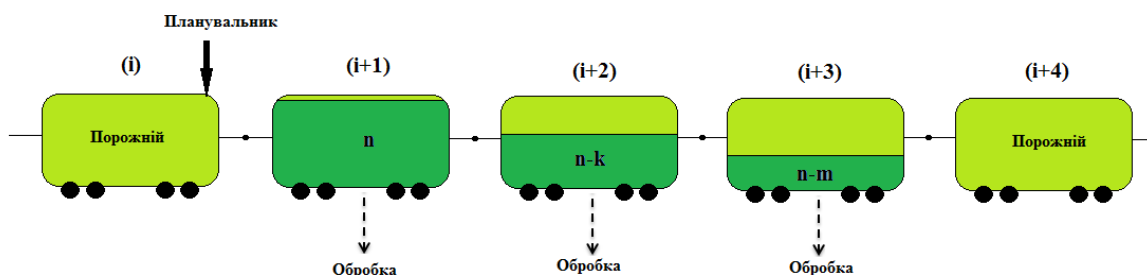


Рис. 4.26 Ефект «нескінченного потягу»

n – максимальна кількість задач для однієї VM

k – кількість задач, які встигли обробитися у $(i+2)$ -й віртуальній машині з початку її наповнення до поточного моменту, моменту завершення подачі заявок у віртуальну машину $(i+1)$.

m – кількість задач, які встигли обробитися у $(i+3)$ -й віртуальній машині з початку її наповнення до поточного моменту.

Метод «нескінченного потягу» є проміжною ланкою між рівнем абстракції та рівнем виконання обчислювальних процедур. Функції контролера, виконання яких не потребує взаємодії з базою даних, можна розглядати як окремі задачі, незалежні одна від одної. Такі функції можна реалізовувати за заданим методом. При цьому окремі задачі можуть бути винесені у хмарне середовище для підвищення ефективності роботи контролера за рахунок збільшення кількості обчислювальних ресурсів.

Основними складовими частинами системи, що розглядається є набір базових функцій контролера та велика кількість прикладних програмних компонентів, які використовують ці функції. Застосування запропонованого методу доцільне в тому випадку, коли система складається з деякого потоку заявок на обслуговування. Розглянемо можливість застосування методу «нескінченного потягу» для кожної з сервісних функцій контролера.

Функція топологій. Функція топологій призначена для зберігання інформації про стан мережі, необхідної для деяких прикладних програмних компонентів. Схема мережі створюється на основі інформації про активність абонентів, якою обмінюються мережеві вузли. Дані, отримані від вузлів

заносяться в базу даних, яка постійно оновлюється. Таким чином, функція топологій взаємодіє з базою даних, тому для неї застосування методу нескінченного потягу не розглядається.

Функція статистики. Всі прикладні програмні компоненти, які працюють з функцією статистики потребують виконання обчислювальних розрахунків. Робота функції статистики включає обробку великого об'єму інформації (даних про стан та активність мережі), виконання обчислювальних операцій з отриманими даними (розрахунок імовірності та середнього часу затримки, довжини черги і т.д.) та прогнозування активності мережі на основі даних отриманих на двох попередніх етапах. Таким чином, дана функція являє собою потік заявок на виконання певних операцій, які не потребують взаємодії з базою даних. Тому для неї є доцільним застосування методу «нескінченного потягу» для збільшення швидкості обробки заявок.

Менеджер комутацій. Дана функція відповідає за побудову логічної архітектури мережі на основі даних функції статистики. Для ідентифікації мережевих елементів використовується фізична топологія мережі. Метод «нескінченного потягу для даної функції не застосовується».

Менеджер потоків. Сервіс організації потоків включає в себе пошук найкоротшого шляху для передачі потоку даних, аналіз інформації про резерв та інші операції керування потоком. Основними завданнями даної функції є ініціація потоку та пошук оптимального шляху. Ці задачі передбачають одночасне обслуговування великої кількості заявок та потребують виконання різних обчислювальних операцій. Тому для оптимізації роботи менеджера потоків можна застосувати метод «нескінченного потягу».

Деякі вендорні компоненти контролеру також можуть бути винесені у хмарне середовище. Це стосується таких платформи-орієнтованих сервісів як координатор Virtual Tenant Network (VTN) та протокол маршрутизації BGP.

VTN координатор є частиною мережевого прикладного програмного компонента сервісного рівня. Він використовує інтерфейс контролера для

побудви віртуальної мережі. Він надає інтерфейс API для прикладних програмних компонентів північного напрямку і знаходить найбільш оптимальні топології мережі. Компонент VTN складається з декількох модулів:

- VTN API – модуль web-інтерфейсу VTN (надає інтерфейс для прикладних програмних компонентів північного напрямку)

- TC – модуль координації
- UPPL – моніторинг/резервування фізичної мережі
- UPLL – моніторинг/резервування віртуальної мережі
- ODC Driver – модуль інтерфейсу контролера

Деякі з цих модулів можуть бути винесені в хмарне середовище.

Протокол маршрутизації BGP використовується контролером SDB в якості протоколу керування платформою для взаємодії між фізичними маршрутизаторами, комутаторами та мережевими сервісами (такими як firewall). Такий підхід дозволяє мережі існувати в різних мульти-вендорних середовищах без необхідності модернізації її інфраструктури. Даний модуль виконує пошук найкоротшого шляху для потоку заявок. Таким чином, кожна заявка повинна до нього звертатись. Тому для цих двох компонентів доцільно реалізувати метод «нескінченного потягу» і перенести деякі їх функціональні модулі у хмарне середовище.

Висновки

Запропонована вперше математичну модель задачі формування вхідного потоку навантаження для ефективного використання ресурсів вузла обслуговування, дозволить розрахувати верхню межу для планування інтенсивності вхідного навантаження із урахуванням бажаної допустимої затримки заявок у системі обслуговування та максимально ефективно використовувати технічні ресурси обслуговування. Обмеження навантаження відповідно до запропонованої моделі дозволило зменшити втрати гібридних ТК сервісів через перевищення черги на 5%.

Запропонована вперше математична модель задачі вибору потужності обслуговуючого пристрою, при якій процес обслуговування утримує на заданому рівні якості, на відміну від існуючих підходів щодо балансування навантаження (коли для забезпечення зростаючого навантаження додається додаткові обслуговуючі пристрої між якими виконується балансування запитів) дозволяє врахувати масштабованість ресурсів вузлів обслуговування побудованих за технологією хмарних обчислень, що мінімізує кількість обслуговуючих пристроїв та втрати часу на балансування на даному рівні представлення. Запропонована модель дозволяє розрахувати параметри системи обслуговування для будь якого вхідного потоку та забезпечити

Запропоновано принцип прогнозування необхідного віртуального ресурсу у хмарі для забезпечення ефективної роботи ГТС, який за рахунок використання даних системи моніторингу для розрахунку конфігурацій системи обслуговування та розкладу їх використання з метою забезпечення показників ефективності, періодичного контролю достатності обраної конфігурації, забезпечить гнучкість та масштабованість системи керування процесом обслуговування гібридних телекомунікаційних сервісів у ГТС, а також ефективне використання віртуальних та фізичних ресурсів обслуговування.

РОЗДІЛ 5

ПРОГНОЗУВАННЯ НЕОБХІДНОГО ВІРТУАЛЬНОГО РЕСУРСУ У ХМАРІ ДЛЯ ЗАБЕЗПЕЧЕННЯ ЕФЕКТИВНОЇ РОБОТИ ГТС

Планування ресурсів гетерогенного телекомунікаційного середовища базується на аналізі великої кількості статистичних даних, дозволяє розрахувати конфігурації обслуговування та план переключення між ними у процесі експлуатації системи керування.

Основними етапами при плануванні ресурсів гетерогенного телекомунікаційного середовища є:

- аналіз даних,
- розрахунок конфігурацій,
- розрахунок плану застосування конфігурацій які забезпечують ефективність використання ресурсів NFV
- динамічний контроль достатності ресурсів відповідно до короткострокової статистики.

Аналіз даних передбачає побудову моделей потоків даних для кожної точки доступу які обслуговуються, моделі статистики навантаження на кожний вузол обслуговування як віртуальний так і фізичний, аналіз карти потоків по мережі.

Розрахунок конфігурацій складається з розрахунку зон обслуговування та розрахінку ресурсів обслуговування відповідно до статистики навантажень на вузли з використанням результатів задачі 1.

Розрахунок конфігурацій зон обслуговування передбачає визначення кожній точці доступу, по якій є статистика навантаження, віртуалізованої (або фізичної) системи ядра оператора зв'язку, яка буде здійснювати обслуговування службових потоків.

Далі за методом побудови розкладу на основі довгострокової статистики визначається графік використання розрахованих конфігурацій системи.

Для забезпечення контролю за перевантаженням у вузлах обслуговування здійснюється моніторинг достатності ресурсів на наступний період часу на основі короткострокової статистики (Рис. 5.1).

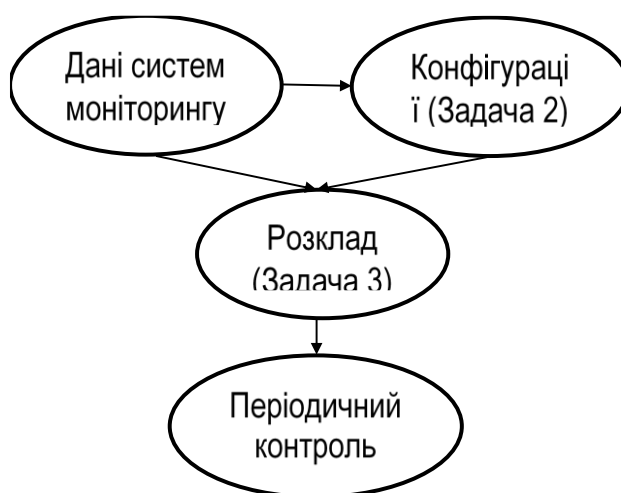


Рис. 5.1 Схема моніторингу достатності ресурсів на основі короткострокової статистики

Мета алгоритму розподілу мережевих ресурсів полягає у тому, щоб їх кількість задовольняла потреби усіх функцій мережі, навіть у моменти пікового навантаження. Завдання полягає у тому, щоб вирішити, в який момент і яка кількість ресурсів повинна надаватись.

Кількість ресурсів, необхідна для кожної мережевої функції, визначається на основі аналітичної моделі мережі. Вхідні дані: вимоги для обслуговування кожного окремого запиту, інтенсивність надходження запитів. Вихідні дані: кількість ресурсів, необхідна кожному функціональному блоку для виконання запитів.

Момент надання ресурсів визначається динамікою навантажень на телекомунікаційну мережу. На зміну навантаження впливають такі фактори, як час доби та сезонні ефекти (довгострокові ефекти), а також незаплановані спалахи активності натовпу (короткострокові ефекти). Довгострокові ефекти,

на відміну від короткострокових, можуть бути передбачені заздалегідь, на основі спостережень і аналізу статистики. Запропонований метод базується на застосуванні двох різних стратегій в умовах змін, які відбуваються в різні періоди часу. Для оцінки і керування навантаженням використовується прогностичне керування ресурсами; для аналізу та виправлення помилок довгострокових прогнозів, а також для реагування на короткострокові спалахнатовпу – реактивне керування ресурсами.

5.1. Система динамічного керування ресурсами мережі

Розглянемо застосування запропонованого методу на прикладі мережі з кількома мережевими функціями. Кожна мережева функція характеризується бажаними вимогами до якості обслуговування – QoS. Передбачається, що вимоги до якості обслуговування визначають терміни цільового часу відповіді. Система керування ресурсами повинна забезпечити таке значення середнього часу відповіді мережевої функції на запит, яке не буде перевищувати бажане цільове значення.

За обслуговування вхідних запитів відповідають відповідні апаратні та програмні ресурси сервера (CPU, NIC і т. д.). Припустимо, що заданий цільовий час відповіді включає в себе кілька значень часу відповіді для кожного конкретного ресурсу. Таким чином, якщо кожен запит на кожному ресурсі не витрачає часу більше, ніж призначене цільове значення, то загальний цільовий час відповіді для сервера буде задоволений. Задача розділення зазначеного значення часу відповіді сервера на значення часу відповіді для конкретного ресурсу виходить за рамки даного дослідження; у роботі передбачається, що такі конкретні для ресурсу значення часу відповіді задані.

Для простоти подання припустимо, що система використовує лише один тип ресурсу. Нехай d_i – цільовий час відповіді мережевої функції i , а T_i – спостережуваний середній час відповіді. Тоді при виділенні ресурсів для

мережевої функції повинна виконуватись умова: $T_i \leq d_i$. Далі буде описано механізм динамічного виділення ресурсів для даної постановки задачі.

Розглянемо архітектуру системи динамічного керування ресурсами і модель масового обслуговування. Проаналізуємо прогностичні та реактивні методи виділення ресурсів.

5.2. Розрахунок конфігурацій системи обслуговування гібридних телекомунікаційних сервісів.

В процесі керування навантаженням необхідно врахувати особливості надходження запитів на обслуговування. Основним джерелом потоків для мобільних мереж зв'язку є мережа доступу. Керування навантаженням для групи вузлів доступу і є предметом дослідження даного методу. Зміна радіусу дії вузла доступу призводить до зміни споживання енергетичних ресурсів точкою доступу, а також до зміни навантаження ну вузлах сусідах. Вибір оптимальних конфігурацій вузлів доступу дозволить забезпечити мінімальну кількість використаних ресурсів та доступність сервісу зв'язку для всіх абонентів, які перебувають на заданій території.

Процес вибір оптимальних конфігурацій складається з двох етапів: розрахунок можливих конфігурацій системи, та вибір оптимальної конфігурації, що відповідає очікуваному навантаженню. Під можливою конфігурацією вузлів обслуговування системи r^k розуміється набір пар значень: $r^k = \{r_i^k, \varphi_i^k\}_{i=1,n}$, де r_i^k – радіус дії і-го вузла доступу в k-й конфігурації системи, φ_i^k – кут (величина сектору) дії і-го вузла доступу в k-й конфігурації системи. У даній моделі покладається, що φ_i^k не залежить від конфігурації що обирається і є постійною величиною $\varphi_i^k = \varphi_i$, однак при необхідності запропонована модель може бути використана для вибору величини сектора дії і-го вузла доступу.

Метод розрахунку можливих конфігурацій системи полягає у вирішенні оптимізаційної задачі:

Дано:

$Node\ i=(x_i, y_i)$ - координати вузлів доступу,

$(l_i^1, l_i^2, Node\ i)$ - сектор дії вузла доступу, характеризується парою напівпрямих, що відсікають сектор кола з центром у $Node\ i$.

$M=\{m_{xy}\}$ - матриця, яка відображає множину точок простору, до яких організовано доступ.

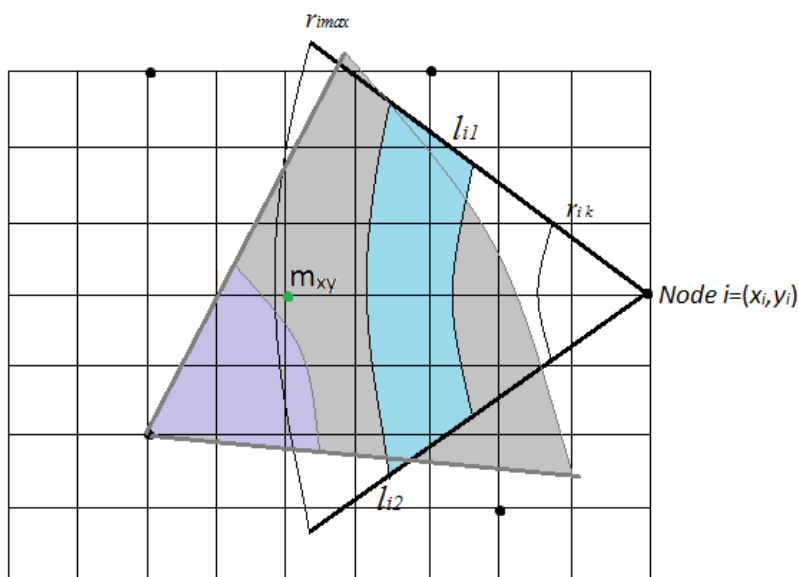


Рис. 5.2 Ілюстрація оптимізаційної задачі

q_d^k - кількість вузлів доступу, з якими має бути зв'язок відповідно до k -ї конфігурації у точок простору з підмножини M_d , експертна характеристика, що залежить від статистики.

$r_{i\ max}$ - максимально можливий радіус дії i -го вузла доступу

M_1, M_2, \dots, M_D ($\bigcup_{d=1, \overline{D}} M_d = M$) - підмножини точок простору які мають спільні статистичні характеристики та вимоги до кількості доступних одночасно базових станцій, $m_{xy}^d \in M_d$ - точка простору із статистичними характеристиками, що відповідають статистичним характеристикам підмножини M_d .

Знайти:

$\vec{r}^k = \{r_i^k\}_{i=1, \overline{n}}$ - набір конфігурацій системи, при яких забезпечується доступ від кожної точки досліджуваної території m_{xy} до вузлів доступу, r_i^k -

радіус дії i -ї точки доступу при конфігурації k . Конфігурації системи відрізняються кількістю одночасно доступних базових станцій для точки m_{xy} , залежно від підмножини. Планування проводиться за умови, що для кожної $m_{xy} \in M$ одночасно доступні не менше двох вузлів доступу.

Запропонований метод:

Введемо позначення $S_i^k = S_i(r_i^k, (l_i^1, l_i^2, \text{Node } i))$ – площа, яку покриває i -й вузол обслуговування при конфігурації k .

Мінімізувати: $r_i^k \rightarrow \min, i = \overline{1, n}$

При обмеженні

$$m_{xy}^d \in \bigcap_{\sum I_i \geq q_d^k} S_i^k, I_i=1, \text{ якщо точка } m_{xy}^d \in S_i^k$$

$$r_i^k \leq r_{i \max}$$

Дану оптимізаційну задачу пропонується вирішувати за допомогою генетичного алгоритму.

Другим етапом в процесі розрахунку конфігурацій системи є визначення конфігурації системи обслуговування базових станцій залежно від змінюваного навантаження та можливостей обслуговування, що надаються вузлом доступу.

Рухомі абонентські термінали зв'язуються з вузлами доступу. Створюючи навантаження на обслуговуючий вузол. Середнє значення навантаження лінійно змінюється. Необхідно розрахувати, на скільки може бути змінена потужність базової станції, в результаті чого зміниться зона покриття. Абоненти, які втратили достатній рівень сигналу з базовою станцією, будуть обслуговуватися сусідніми вузлами доступу. Припускаємо, що заявки абонентів рівномірно перерозподіляються між базовими станціями-сусідами. Розрахувати як може бути змінена потужність базової станції, щоб показники якості обслуговування залишилися на заданому рівні, а кількість спожитих енергоресурсів була мінімальною.

Визначення сусідів для вузла обслуговування є складною задачею оскільки, за умовами методу навантаження, що створюється точками m_{xy}

рівномірно перерозподіляється між базовими станціями сусідами, тому важливо визначити якою мірою мають накладатися зони дії вузлів доступу для забезпечення якісного обслуговування абонентів. Для визначення оптимальної площі перетину зон дій базових станцій для забезпечення ефективного перерозподілу було проведено моделювання.

5.2.1. Вибір найкращої конфігурації, що забезпечує енергоефективну роботу ГТС

Вхідні дані.

G^k – граф вузлів обслуговування.

$\{g_{ij}^k\}_{i,j=\overline{1,n}}$ – матриця зв'язності графа G , n – кількість вузлів графа, $g_{ij}=1$, якщо вузлами i та j , якщо вони є вузлами-сусідами у конфігурації k , $g_{ij}=0$ – інакше. Вузли i та j є вузлами-сусідами, якщо площа перетину зон дій вузлів i та j не менше $\frac{1}{2}$ меншої площі зони дії цих вузлів: $S_i^k \cap S_j^k \geq \min(S_i^k/2, S_j^k/2)$.

μ_i – інтенсивність обслуговування заявки у вузлі i ($i = \overline{1,n}$), в рамках даного методу не змінюється.

λ_i – інтенсивність вхідного потоку заявок у вузол i до перерозподілу ($i = \overline{1,n}$),

$f_i(r_i, \rho_i)$ – значення функції енергоспоживання вузла i ($i = \overline{1,n}$), при завантаженості вузла $\rho_i = \lambda_i / \mu_i$, та радіусі дії r_i , функція задана технічними характеристиками вузла обслуговування або отримана експериментально.

$\lambda_{i \max}$ – інтенсивність вхідного потоку заявок, яка є оптимальною для вузла i ($i = \overline{1,n}$), яка розрахована за методом формування вхідного потоку навантаження для ефективного використання ресурсів обслуговування, описаного вище.

$\{\vec{r}^k\}$ – послідовність можливих значень радіусу дії вузла i ($i = \overline{1,n}$). Зміна навантаження на вузол обслуговування відбувається за рахунок зміни

потужності випромінювання базової станції (БС), що призводить до зміни радіусу дії БС r_i , відповідно до зменшення кількості абонентів, які зв'язуються з БС та створюють навантаження. Припускається, що абоненти рівномірно розташовані, на території, що покривається БС.

Вихідні дані.

k – оптимальна конфігурація для вузлів доступу, що забезпечить мінімізацію сумарного енергоспоживання мережі обслуговуючих вузлів (W_{\min}).

Запропонований метод.

Цільова функція запропонованого методу: $\sum_{i=1}^n f_i(r_i^k, \lambda_i^*) \rightarrow \min$

За умов:

$$\lambda_i^* \leq \lambda_{i \max} \quad (5.1)$$

1. При зменшенні інтенсивності вхідного потоку у вузлі i на $(\lambda_i - \lambda_i^*)$ інтенсивність вузлів-сусідів збільшується на $\lambda_{ij} = (\lambda_i - \lambda_i^*) / \sum_{j=1}^n g_{ij}$.

Для розв'язку даної задачі я пропоную застосовувати метод динамічного програмування. Відповідно до якого, на кожному етапі послідовно розраховується енергоспоживання вузла для кожної конфігурації вузла доступу.

Введемо позначення $\overrightarrow{\lambda^{xk}}$ – вектор значень інтенсивності вхідного навантаження на вузли мережі обслуговування розраховані на етапі x , що відповідає радіусу дії вузла обслуговування x у k -й конфігурації – r_x^k . Координата вектора $\overrightarrow{\lambda^{xk}}$ – λ_i^{xk} – це інтенсивність надходження заявок у вузол i ($i = \overline{1, n}$), яка розрахована на етапі x та відповідає радіусу дії вузла обслуговування x – r_x^k , ($k = \overline{1, K}$). Дана величина збільшується при перерозподілі абонентів у вузлах $j \in \{j | g_{ij} = 1\}$, та зменшується при перерозподілі абонентів у вузлі i .

Відповідно до припущення про рівномірність розташування абонентів у зоні покриття вузла зменшення радіусу дії з r_i до r_i^k , зменшить λ_i^x , нове

значення $\lambda_i^x = \frac{r_i^k \lambda_i^x}{r_i}$, перерозподілене навантаження складе величину $\lambda_{ij} = \left(\lambda_i^x - \frac{r_i^k \lambda_i^x}{r_i} \right) / \sum_{j=1}^n g_{ij}$.

Введемо позначення:

$W_n(r_n^k, \overrightarrow{\lambda^{nk}}) = f_n(r_n^k, \lambda_n^x)$ – енергоспоживання вузла n , якщо радіус його дії r_n^k , а інтенсивність вхідного навантаження λ_n^x ;

$W_z(r_z^k, \overrightarrow{\lambda^{zk}})$ – сумарне енергоспоживання вузлів $z, z+1, \dots, n$, якщо радіус дії вузла $z - r_z^k$, а інтенсивність вхідного навантаження – λ_z^x ;

$W_1(r_1^k, \overrightarrow{\lambda^{1k}})$ – сумарне енергоспоживання всіх системи обслуговування, радіус дії вузла $1 - r_1^k$, а інтенсивність вхідного навантаження – λ_z^x .

Алгоритм пошуку оптимального значення:

Шаг 1. Для вузла n , для кожної конфігурації r_n^k , розрахувати вектор $\overrightarrow{\lambda^{nk}}$ таким чином, щоб

$$\overrightarrow{\lambda^{nk}} = \begin{cases} \lambda_n^{nk} = \frac{r_n^k \lambda_n}{r_i} \\ \lambda_i^{nk} = \frac{\left(\lambda_n - \frac{r_i^k \lambda_n}{r_i} \right)}{\sum_{j=1}^n g_{nj}}, & \text{для всіх } \{i | g_{in} = 1\} \\ \lambda_i^{nk} = \lambda_i, & \text{для всіх } \{i | g_{in} = 0\} \end{cases}$$

Якщо існує таке i для якого не виконується нерівність $\lambda_i^{nk} \leq \lambda_{i \max}$, тоді відповідна конфігурація з розгляду. Таким чином, $R_n = \{k | \lambda_i^{nk} \leq \lambda_{i \max}\}$ – множина конфігурацій, які розглядаються на наступних шагах.

Для всіх конфігурацій r_n^k вузла n розрахувати, $W_n(r_n^k, \overrightarrow{\lambda^{nk}})$.

Шаг z . Для вузла z , для кожного $k \in R_z$ розрахувати вектор $\overrightarrow{\lambda^{zk}}$ таким чином, щоб

$$\overrightarrow{\lambda^{zk}} = \begin{cases} \lambda_z^{zk} = \frac{r_z^k \lambda_z^{(z+1)k}}{r_z} \\ \lambda_i^{zk} = \frac{\left(\lambda_n^{(z+1)k} - \frac{r_i^k \lambda_n^{(z+1)k}}{r_i} \right)}{\sum_{j=1}^n g_{nj}}, & \text{для всіх } \{i | g_{iz} = 1\} \\ \lambda_i^{zk} = \lambda_i^{(z+1)k}, & \text{для всіх } \{i | g_{iz} = 0\} \end{cases}$$

Перевіряється виконання умови $\lambda_i^* \leq \lambda_{i \max}$

(5.1), та формується множина $R_z = \{k | \lambda_i^{zk} \leq \lambda_{i \max}\}$

Для всіх $k \in R_z$ вузла z розрахувати

$$W_z(r_z^k, \overrightarrow{\lambda^{zk}}) = f(r_z^k, \lambda_z^{zk}) + W_{z+1}(r_{z+1}^k, \overrightarrow{\lambda^{(z+1)k}}) \quad (5.2)$$

Шаг n . Для вузла 1, для кожного $k \in R_z$ розрахувати вектор $\overrightarrow{\lambda^{1k}}$, за формулами (5.2) при $z=1$, знайти множини $R_1 = \{k | \lambda_i^{1k} \leq \lambda_{i \max}\}$.

Для всіх $k \in R_1$ за формулою розрахувати, $W_1(r_1^k, \overrightarrow{\lambda^{1k}})$.

Мінімальне значення енергоспоживання системи обслуговуючих вузлів знаходиться за формулою (5.3):

$$W_{\min} = \min_k \left(W_1(r_1^k, \overrightarrow{\lambda^{1k}}) \right) \quad (5.3)$$

Оптимальна конфігурація знаходиться як $k = \arg \min W_{\min}$.

Запропонований метод дозволяє розрахувати допустимі конфігурації вузлів доступу, що забезпечують доступність сервісу зв'язку у будь-якій точці території, яка обслуговується групою вузлів доступу, та визначити який набір конфігурацій забезпечить мінімальне сумарне споживання енергетичних ресурсів вузлами доступу.

5.3. Метод складання розкладу використання конфігурацій

Для забезпечення безперебійної роботи серверу тарифікації необхідно розробити комплекс засобів, які враховують не тільки поточну ситуацію, але й статистику набрану за тривалий період часу. Це дозволить планувати ресурси, розробляти методи балансування навантаження, тощо.

Для рішення даної задачі необхідно провести аналіз статистики протягом тривалого періоду. Наприклад, аналіз статистики за понеділок протягом 15-20 тижнів. Результатом таких статистичних даних стане матриця де, для кожного малого інтервалу часу, наприклад, для часу 8:00:00-8:00:01, такий інтервал позначатимемо значенням закінчення інтервалу: 8:00:01. В матриці будуть зібрані дані про кількість заявок (навантаженні), яке потребувало обслуговування на всіх доступних серверах для яких складається даний розклад. Розглядатимемо групу серверів як єдину систему з одним входом, задача розподілу вхідного навантаження між доступними серверами не розглядається. Необхідно знайти скільки серверів з відомими технічними характеристиками повинно бути залучено до обробки вхідного навантаження для забезпечення мінімуму втрат через перевантаження серверу.

Першим етапом методу є розбиття загального періоду часу на відрізки, які характеризуються однаковою динамікою зміни навантаження. Оцінка динаміки зміни навантаження буде проводитися на основі аналізу прямої, яка апроксимує статистичні дані.

Метод побудови розкладу вибору конфігурацій на основі довгострокової статистики базується на великій кількості статистичних даних які збираються протягом тривалого часу (наприклад протягом року). Ці дані групуються по різним дням тижня. Аналізується довгострокова статистика навантаження, що створюється групою вузлів доступу, якщо даний метод застосовується для планування конфігурацій груп вузлів обслуговування, або навантаження на вузол обслуговування, якщо метод застосовується для планування конфігурацій розподілу ресурсів вузлів.

Метод вирішує завдання визначення порядку застосування конфігурацій ресурсів обслуговування, а саме моментів переключення (рис. 5.3). Основна наукова ідея методу полягає у ітераційному визначенні конфігурацій для різних інтервалів часу та перевірки нерівності Чебишова

для забезпечення вимог щодо достатності ресурсів обслуговування протягом всієї часової прямої.

Вхідні дані:

- λ_k – максимальна інтенсивність обслуговування, яка відповідає k -й конфігурації;
- $\vec{\lambda}_i$ – набір векторів статистичних значень кількості заявок, $t \in [0, T]$;
- $\vec{\lambda}_t = (\lambda_t^1, \lambda_t^2, \dots, \lambda_t^{50})$ – це набір значень кількості заявок, які надійшли у i -й момент часу, за останні 50 тижнів.

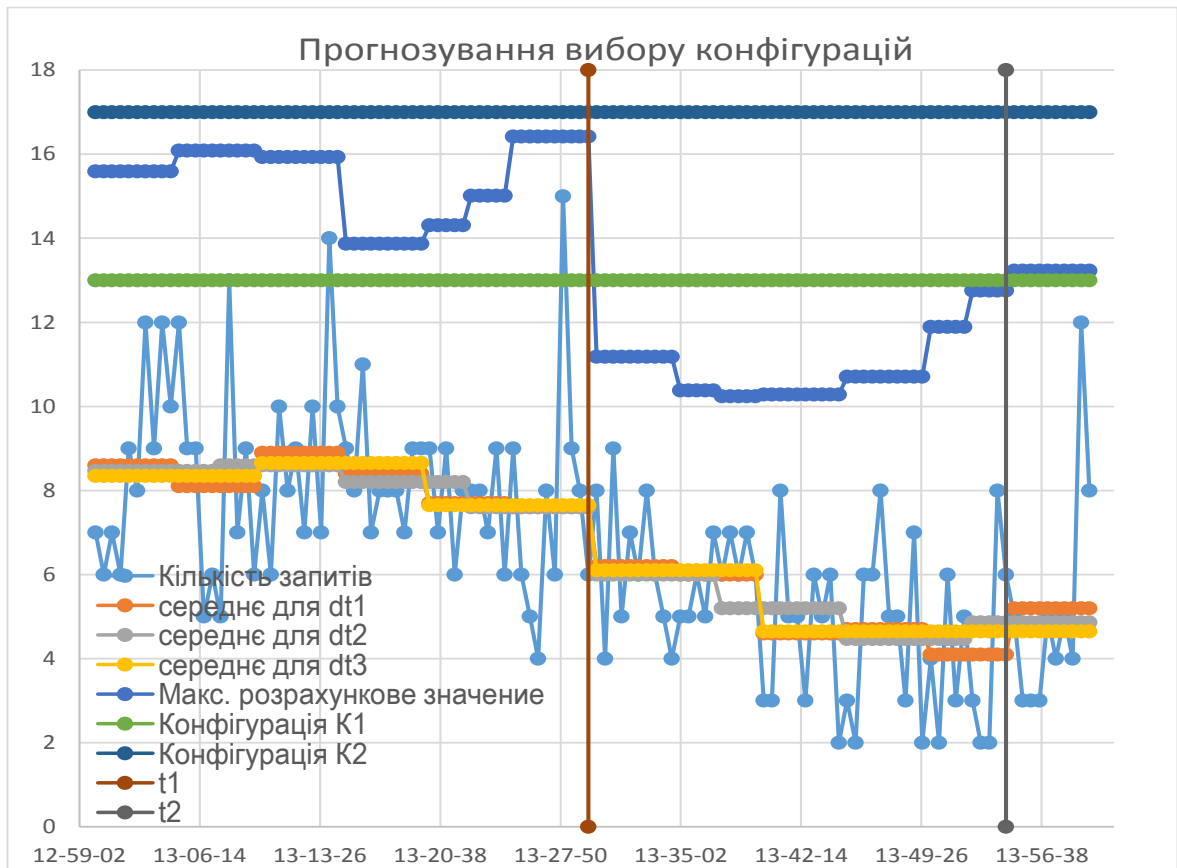


Рис. 5.3. Метод побудови розкладу вибору конфігурацій на основі довгострокової статистики

Вихідні дані:

- (k, t) , де $t \in [0, T]$ – рекомендовані моменти зміни інтенсивності обслуговування, k – рекомендована конфігурація системи, на яку слід перейти в момент часу t .

Алгоритм методу:

Етап 1. Розбити часову вісь на інтервали часу тривалістю Δt ;

Етап 2. Для кожного інтервалу $\forall i = \overline{1, \frac{24}{\Delta t}}$ знайти середнє значення $M\lambda$ та дисперсію σ^2 . Для аналізу береться множина даних зібраних за тривалий час обслуговування;

Етап 3. Для кожного інтервалу знайти мінімальну конфігурацію, для якої виконується нерівність:

$$M\lambda + 3\sigma^2 \leq \lambda_k$$

Створити тимчасову множину $Kt = \{Kt_i\}$, $Kt_i = (k, i)$, $i = \overline{1, T}$

Етап 4. Створити постійну множину Kt' , якщо $Kt' = \emptyset$, то $Kt' := Kt$, Етап 5; інакше попарно порівняти елементи множин Kt'_i та Kt_i , якщо $\forall Kt'_i$ виконується $Kt'_i = Kt_i$, то Етап 6, інакше $Kt'_i = \max(Kt'_i, Kt_i)$, Етап 5;

Етап 5. Змінити значення Δt , перейти на Етап 2;

Етап 6. На основі Kt'_i визначити (k, t) , де $t \in [0, T]$ – рекомендовані моменти зміни інтенсивності обслуговування, k – рекомендована конфігурація системи, на яку слід перейти в момент часу t .

Результатом роботи даного алгоритму є матриця, перший рядок якої відображає моменти часу в які необхідно проводити операцію включення або виключення серверу, другий рядок означає кількість активних серверів які повинні обслуговувати вхідне навантаження у наступний період часу, третій рядок означає процес який має бути ініційований у відповідний момент часу: 1 – включення одного з серверів, 0 – виключення одного з серверів.

Слід зауважити, що робота даного алгоритму є лише складовою частиною процесу технічного забезпечення системи онлайн тарифікації. Оскільки крім статичного розкладу включення серверів потребує розгляду питання балансування навантаження між серверами.

5.4. Оцінка ефективності статистичного методу розподілу кількості хмарних ресурсів.

Як наводилося на початку статті, спектр проблем, які можуть вирішуватися за допомогою запропонованого методу та алгоритмів досить широкий. Основною ознакою систем, для яких може бути застосований запропонований метод, є виконання великої кількості процедур, ініціатором яких є люди або інші програми. Виконання процедур здійснюється за допомогою програмного забезпечення серверу, при цьому використовують технічні ресурси системи.

На сьогоднішній день проблема перевантаження вирішується за рахунок не допуску в систему надлишкової кількості запитів на тарифікацію, тобто якщо сервер є перевантаженим, то поступає сигнал на керуючий пристрій і тимчасово заявки на тарифікацію не приймаються. При цьому одночасно задіяні (знаходяться у режимі очікування) всі ресурси серверу, завантаження обладнання на 20-30% є нормальним в процесі обслуговування.

Був проведений експеримент роботи імітаційної моделі в двох режимах:

1. На обслуговування виділявся обмежений технічний ресурс, у разі перевантаження заявки відкидалися (Режим без застосування запропонованих методів).

2. Обслуговування заявок може проводитися від одного до трьох аналогічних серверів з обмеженими технічними ресурсами, на основі статистичних вибірки від оператора зв'язку був сформований розклад включення серверів відповідно до алгоритму пошуку моментів переключення, далі було згенеровано вхідний потік максимально наближений до реального. Включення серверів відбувалося за розкладом, а

за методом включення додаткового технічного обладнання кожну три хвилини відбувалась перевірка достатності ресурсів для обслуговування вхідного потоку.

За результатами роботи імітаційної моделі було проведено дослідження використання електричної енергії, яке показало скорочення споживання енергії на 60% (Рис. 5.3 Порівняння використання електричної енергії).

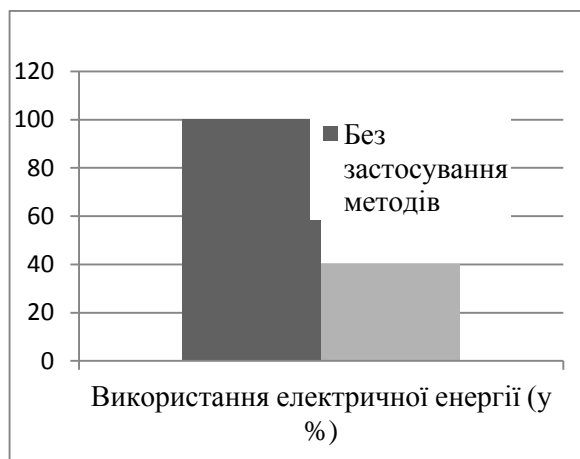


Рис. 5.3 Порівняння використання електричної енергії

Розглянуті проблеми організації роботи високонавантажених систем. На прикладі роботи системи онлайн тарифікації наведено принцип організації процесу обслуговування. Запропоновано метод складання розкладу включення серверів забезпечує довгостроковий розклад включення серверів, побудований на даних довгострокової статистики, дозволить спланувати роботи технічних засобів на тривалий час. Проведене моделювання показало, що скорочується споживання електричної енергії на 60%

5.5. Контроль достатності зарезервованих ресурсів

На сьогоднішній день спостерігається перехід до використання глобальних інформаційних ресурсів. Що передбачає необхідність організації процесу одночасного обслуговування великої кількості користувачів. Список сервісів, які обслуговують мільйони користувачів все збільшується, починаючи від соціальних мереж та пошукових машин і закінчуючи системами онлайн тарифікації крупних операторів зв'язку. Основними

вимогами до технічних ресурсів, що обслуговують подібні системи є забезпечення доступності сервісу та безперебійної роботи.

Організація процесу надання послуг великій кількості користувачів, яка оцінюється в мільйони, має спільні риси. Так все більше компаній користуються послугами облачних сервісів, які забезпечують технічними ресурсами та можуть надавати інфраструктури як сервіс.

На відміну від моделі зберігання даних на власних виділених серверах, у випадку використання віртуальних серверів, їх структура і кількість в загальному випадку не видна користувачеві. Вся інформація зберігається і обробляється у хмарі, яка являє собою, з точки зору клієнта, один великий віртуальний сервер. На сьогоднішній день велика кількість компаній надає таку послугу як оренда сервера у хмарі.

Основні переваги віртуальних серверів:

- *Гнучкість ресурсів.* Оренда сервера у хмарі дозволяє забезпечити його високу масштабованість. Такий сервер легко налаштовується під збільшення навантаження, наприклад, можна легко додати оперативної пам'яті або дискового простору. Так само легко можна зменшити дані параметри віртуальної системи. Сервер стає «резиновим» у відношенні своїх ресурсів.

- *Швидкодія.* Віртуальний сервер працює значно швидше ніж звичайний, а також знижуються часові затрати на впровадження і оперативний перерозподіл ресурсів. Висока швидкість розгортання системи.

- *Безпека.* На віртуальному сервері клієнт має можливість власноручно контролювати користувачів, процеси, а також використовувати власну політику безпеки. Підвищення рівня безпеки Cloud Server відбувається також за рахунок зведення до мінімуму «людського фактору».

- *Мобільність.* Доступ до серверу можна отримати з будь-якої точки земної кулі.

Для забезпечення роботи сервісів використовується модель функціонування зображена на Рис. 5.4.

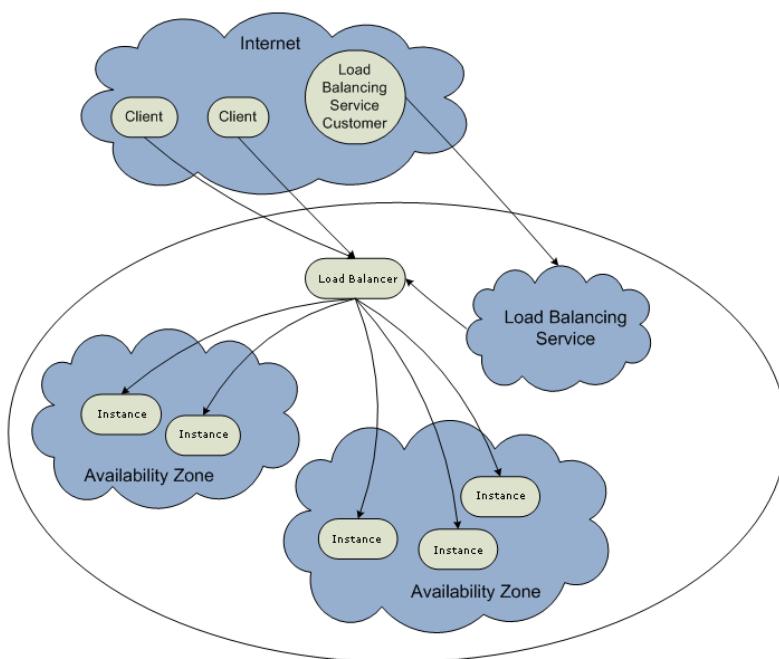


Рис. 5.4 Організація процесу розподілу задач високонавантажених системи в Cloud.

Розглянемо процес обслуговування великої кількості абонентів на прикладі системи онлайн тарифікації опаратора зв'язку. Використання віртуальних серверів для обслуговування і тарифікації послуг може значно спростити процес розподілу ресурсів на сервері тарифікації, забезпечить його масштабованість, підвищить безпеку і відказостійкість. Віртуальний сервер може бути використаний трьома способами. По-перше, він може бути додатком до основного фізичного сервера і використовуватись коли виникає потреба у збільшенні кількості ресурсів при збільшенні навантаження. В такій конфігурації вся логіка серверу реалізована у хмарі. В другому варіанті все обладнання для обслуговування і тарифікації викликів може розміщуватись у хмарі. В такому випадку фізичний сервер взагалі не потрібен і процес обслуговування запитів буде повністю проходити на віртуальному сервері. В третьому випадку, який розробляється, у хмарі можуть знаходитися окремі ресурсозатратні підсистеми або частини підсистем, організація взаємодії яких є окремою мережевою задачею.

Наведені вище способи значно розширюють можливості технічного забезпечення системи онлайн тарифікації. При цьому для ефективного

планування необхідної кількості технічних ресурсів необхідно розробити методи, що дозволяють планувати кількість необхідних ресурсів. В (54) біло запропоновано динамічний метод оцінки достатності підключених ресурсів у поточний момент часу. В даній статті пропонується довгостроковий метод, який забезпечує безперебійну та економічну роботу системи тарифікації. Економічність полягає в тому, що відповідно до статистичних даних про вхідне навантаження, відбувається включення потрібної кількості серверів, які забезпечують роботу системи тарифікації або його підсистем. Фактично такий розклад визначає між скількома обслуговуючими одиницями (Рис. 5.4) буде здійснюватися балансування навантаження.

5.5.1. Алгоритм пошуку інтервалу контролю достатності ресурсів

Динамічне керування ресурсами потребує використання на кожному сервері трьох компонентів:

- 1) модуль моніторингу стану мережі – вимірює навантаження і показники продуктивності мережевих функцій;
- 2) модуль прогнозування – використовує дані модуля моніторингу для оцінки характеристик навантаження в найближчому майбутньому;
- 3) модуль розподілу ресурсів – використовує оцінки модуля прогнозування для виділення необхідної кількості ресурсів мережевими функціями.

Розглянемо систему гібридного керування мережевими ресурсами. Незважаючи на те, що навантаження мають передбачувані шаблони, можливі значні відхилення від існуючих моделей, так як навантаження датацентрів мають пульсуючий характер. Надлишкові заявки або раптові сплески запитів можуть привести до погіршення обслуговування. Крім того, існують вартість та ризик пов'язані зі змінами у системі, і, таким чином, прагнуть уникнути частих змін виділення ресурсів.

На основі вищезазначених спостережень пропонується гібридне рішення системи керування виділенням ресурсів, яка включає в себе прогностичну та реактивну складову. Основна ідея підходу полягає у фіксуванні періодичних і стійких моделей навантаження з урахуванням історичних даних (базового навантаження). Після цього відбувається проактивне і прогностичне керування ресурсами мережі з метою обслуговування навантаження. Різниця фактичного і передбаченого навантаження обробляється реактивним способом. Крім того, розроблена система враховує витрати і ризики, пов'язані з керуванням ресурсами.

Архітектура системи гібридного керування зображена на Рис. 5.5.

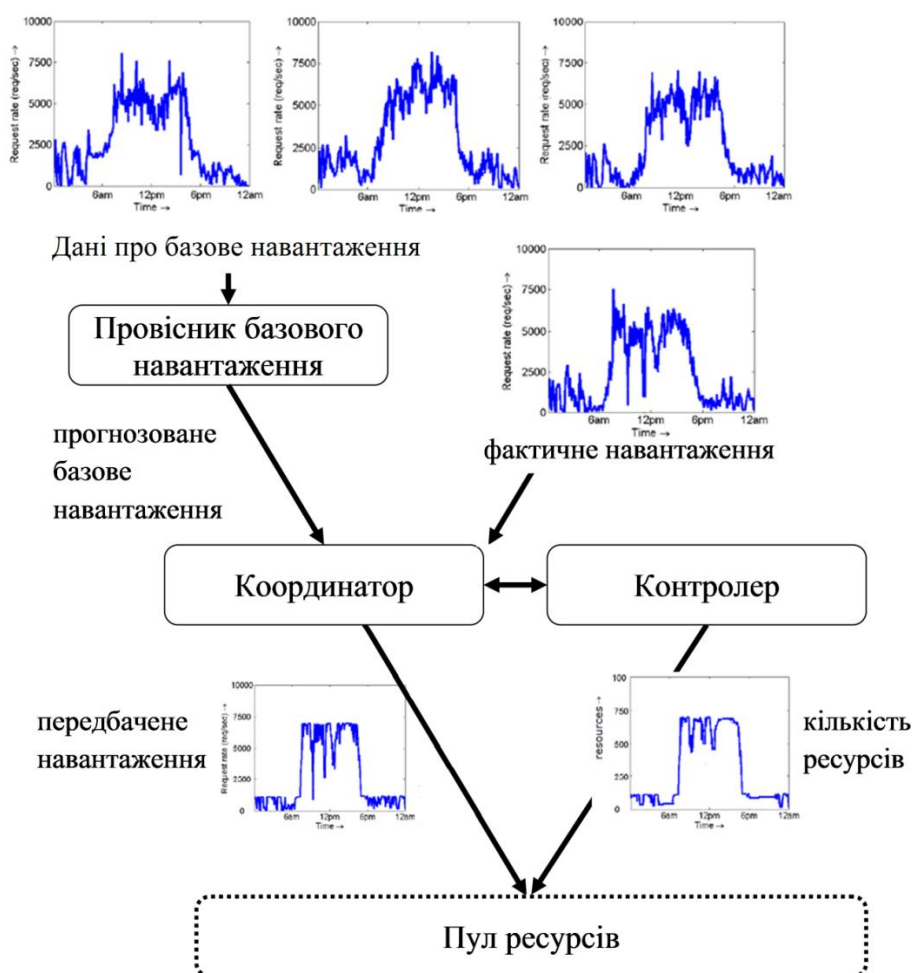


Рис. 5.5 Архітектура системи гібридного керування ресурсами

Далі розглянемо більш детально принцип роботи модуля моніторингу, який відповідає за виконання вимірів в реальному режимі.

5.6. Модуль моніторингу та вимірювання величин

Основна задача даного модулю полягає у виконанні вимірів різних метрик системи у реальному часі. Ці показники необхідні для оцінки параметрів моделі і отримання характеристик навантаження.

При використанні традиційних методів моніторингу стану мережевих ресурсів, збільшується кількість надлишкової службової інформації, що негативно впливає на ефективність роботи всієї мережі через збільшення завантаженості каналів. Для вирішення цієї проблеми і забезпечення керування системою, побудованою на основі технології NFV, необхідно модернізувати існуючі механізми моніторингу. Для цього пропонується використовувати метод динамічного моніторингу.

Суть методу динамічного моніторингу полягає у постійній зміні інтенсивності виконання вимірів в залежності від точності їх прогнозування, тобто від різниці між передбаченим і фактичним значенням навантаження. Прогнозоване значення надходить на елемент мережі на основі даних стану елемента, зібраних у реальному часі. При збільшенні відхилення, інтервал між опитування мережевих вузлів скорочується, тобто система переходить у стан пильного моніторингу.

Реалізація такого інтелектуального моніторингу відбувається шляхом встановлення на координатор головного прикладного програмного компонента моніторингу та підконтрольних їй агентів на кожен мережевий блок. Програма на координаторі відповідає за здійснення аналізу даних моніторингу, а підконтрольні агенти збирають відповідну інформацію, модифікують її та відправляють на координатор.

Зміна інтенсивності моніторингу сегментів мережі в залежності від похибки прогнозування дозволяє усунути надлишкові процеси моніторингу та обробки даних. Використання такого підходу надає можливість підвищити гнучкість керування мережевими елементами, а отже більш ефективно використовувати ресурси пристроїв рівня керування. Володіння актуальною

службовою інформацією також дозволить керувати навантаженням мережі і зменшити витрати на електроенергію, продовживши тривалість роботи мережевих елементів.

Зміна частоти виконання моніторингу описує формула (5.4)

$$I(t) = I_{base} - K \cdot \sum_{j=t-h}^{t-1} \frac{(\lambda_{obs}(j) - \lambda_{pred}(j))^+}{h} \cdot I_{base} \quad (5.4)$$

Де I - інтервал моніторингу;

I_{base} – базове значення інтервалу;

K – константа нормалізації;

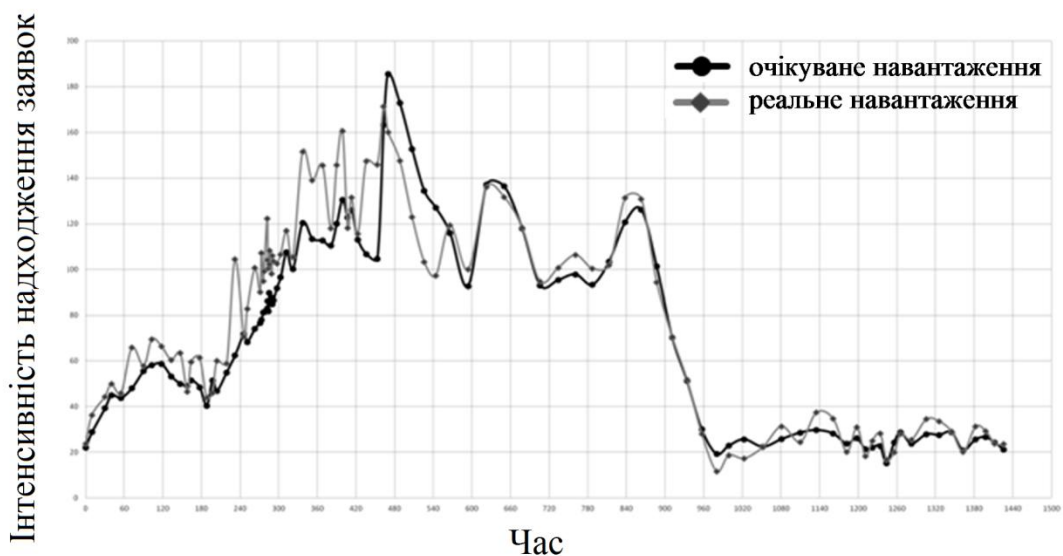
$\lambda_{obs}(t)$ – реальна інтенсивність надходження навантаження протягом інтервалу t ;

$\lambda_{pred}(t)$ – передбачена інтенсивність надходження навантаження під час інтервалу t .

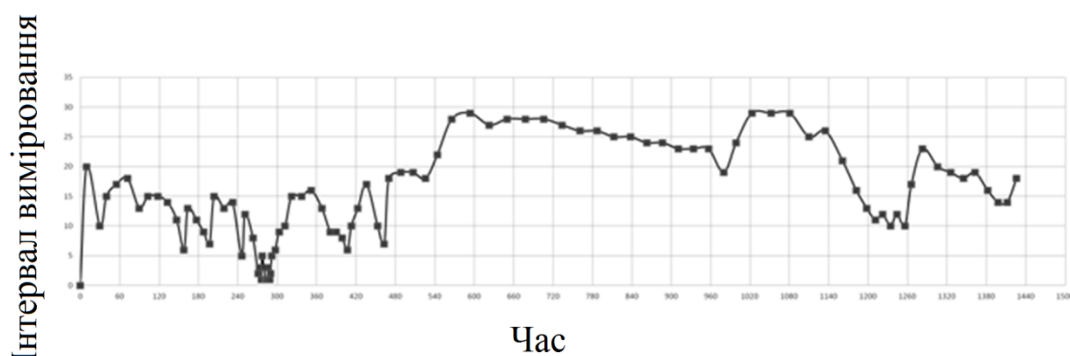
Адаптація частоти здійснення вимірів до відхилення реального навантаження від очікуваного на мережевому елементі зображена на Рис. 5.6.

Під час виконання вимірювань враховуються такі часові інтервали, як історія вимірювання (H) і вікно адаптації (W) (Рис. 5.7).

Інтервал вимірювання – це інтервал, на якому розташовані параметри, що враховуються при виконанні вимірювання. Модуль моніторингу аналізує кількість вхідних запитів на кожному інтервалі і записує отримане значення. Вибір інтервалу вимірювання залежить від бажаного відгуку системи. Якщо потрібно, щоб система реагувала на зміну навантаження на дрібній часовій шкалі, потрібно вибрати невелике значення інтервалу вимірювання, у іншому випадку – велике значення.



(а)



(б)

Рис. 5.6 Адаптація частоти моніторингу: а) очікуване та реальне навантаження мережевого елемента; б) відхилення реального навантаження від очікуваного

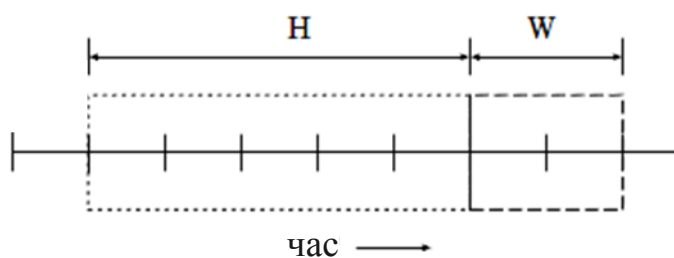


Рис. 5.7 Часові інтервали вимірювань

Історія вимірювання – це послідовність зареєстрованих значень для кожного необхідного параметра. Множина значень історії, яка включає в себе H останніх значень для кожного параметра, зберігається на модулі

моніторингу. Передбачення майбутніх значень параметрів відбувається на основі цього базису.

Вікно адаптації – часовий інтервал між двома послідовними запусками алгоритму адаптації. Значення минулих вимірювань використовують для прогнозування навантаження інтервалу часу W , а система адаптується для цього інтервалу. Як можна буде побачити далі, запропонований опис моделі масового обслуговування у часовій області розглядає період часу рівний вікну адаптації для оцінки середнього часу відповіді T_i мережевої функції, і ця модель оновлюється після інтервалу W або коли досягається заданий поріг.

Визначимо і дискретизуємо шаблони прогнозованого навантаження. Основна задача полягає в тому, щоб представити денний шаблон в навантаженні, дискретизуючи його запити у послідовні, непересічні часові інтервали з єдиним репрезентативним значенням вимоги в кожному інтервалі. Для цього використаємо алгоритм, який дозволяє знайти невелику кількість часових інтервалів, для яких відхилення від фактичної вимоги зводиться до мінімуму. Важливо, щоб таких інтервалів було небагато, так як збільшення їх кількості призведе до більш частих змін у виділенні ресурсів, а отже, більш високих витрат. Дамо формальне визначення дискретизації.

Визначення дискретизації навантаження:

Якщо існує деякий числовий ряд X на області $[v, \tau]$, то числовий ряд Y , визначений на тій самій області, є характеристикою навантаження X якщо область $[v, \tau]$ можна розділити на m послідовних непересічних часових інтервалів: $\{[v, \tau_1], [\tau_1, \tau_2], \dots, [\tau_{m-1}, \tau]\}$, так що $X(j) = r_i$, для всіх j у i -му інтервалі, $[\tau_{i-1}, \tau_i]$.

Будь-який часовий ряд X сам по собі є дискретизацією. Для випадку, що розглядається встановимо $v=0$; τ – період навантаження. На основі аналізу періодичності припустимо, що період навантаження дорівнює 24 години.

Ідея дискретизації навантаження є подвійною. По-перше, необхідно точно відображати вимоги. Щоб досягти цього, репрезентативні значення r_i , для кожного інтервалу $[\tau_{i-1}, \tau_i]$ повинні бути максимально близькими до фактичних значень часового ряду на інтервалі $[\tau_{i-1}, \tau_i]$. По-друге, для виділення ІТ-ресурсів необхідні кошти. Тому занадто велика кількість інтервалів, а отже, і змін в системі, є непрактичною і може призвести до великої кількості проблем: втрати продуктивності, нестабільності системи, зносу ресурсів і т. д. Щоб уникнути цього, необхідно зменшити помилку, яку внесла дискретизація, а також кількість інтервалів дискретизації. Для мінімізації цих двох величин введемо наступний вираз:

$$\sum_{i=1}^m \left[\sum_{\tau=\tau_{i-1}}^{\tau_i} u(r_i - X(\tau)) \right] + f(m) \rightarrow \min \quad (0.5)$$

Даний описує цільову функцію, яку потрібно мінімізувати. У ньому:

X – часовий ряд;

$f(m)$ – функція вартості кількості змін або інтервалів m .

Для виконання виразу (5.8), необхідно одночасно мінімізувати помилку репрезентації навантаження і кількість змін. Для деяких випадків доцільно віддати перевагу мінімізації квадрату кількості змін, або іншій функції кількості змін. Вибором найкращої функції вартості займаються мережеві адміністратори. При цьому враховуються конкретні потреби мережі, яка знаходиться в керуванні.

Для випадку, що розглядається встановимо значення функції вартості $f(m) = l \times m$, де l – константа нормалізації. Метою цільової функції є мінімізація нормалізованої кількості змін і помилки репрезентації. Функція вартості помилки репрезентації необхідна для кількісної оцінки похибки визначення навантаження. В загальному випадку кількість змін і помилка репрезентації можуть бути сформульовані у вигляді функцій корисності.

В якості функції кількості змін може виступати вартість електроенергії, або витрати, пов'язані зі зносом не перезавантаження, які базуються на

номінальному часі роботи на відмову і вартості заміни диска. Оптимальна величина інтервалу визначається ітераційно. Для цього задамо різні значення кількості інтервалів і обчислимо значення виразу. З отриманих значень вибираємо мінімальне на кожному інтервалі $r_i = \max_{\tau \in (\tau_{i-1}, \tau_i)} X(\tau)$.

На Рис. 5.8 представлено приклади значень числового ряду. Помилка репрезентації для інтервалу 10 хвилин складає 7%, для випадку 60 хвилин – 19%.

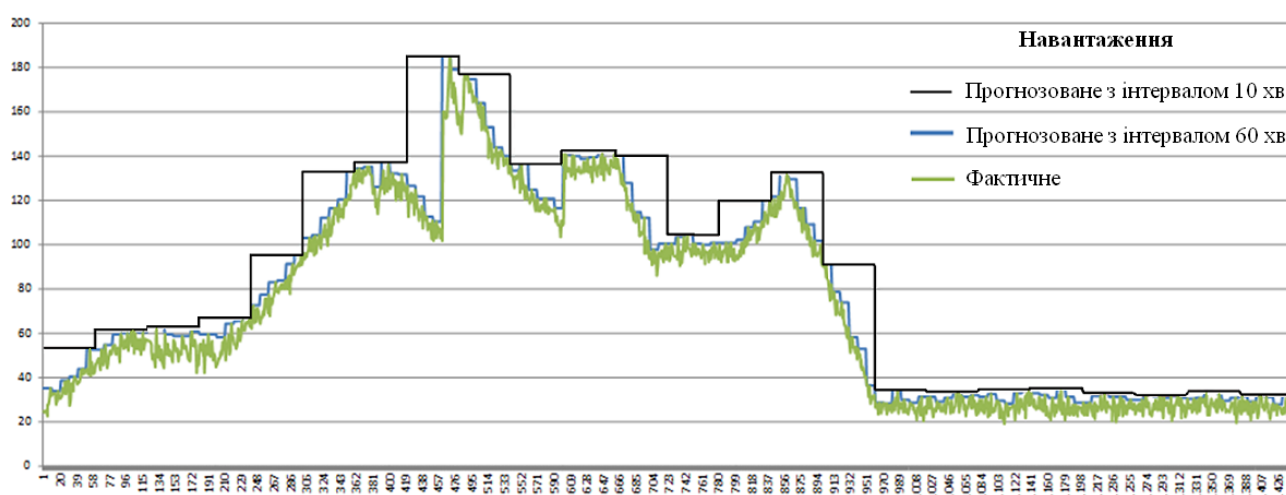


Рис. 5.8 Значення навантаження в залежності від величини інтервалу адаптації

Псевдокод даного алгоритму буде мати наступний вигляд:

```

interval ← maxinterval
iinterval ← 0
startpoint ← beginning
while interval > 0
    startpoint ← 0
    endpoint ← startpoint + interval
    while startpoint < endpoint
        maxvalue ← max[Load(startpoint)... Load(endpoint)]
        for j ∈ startpoint... endpoint
            BaseLoadinterval(j) ← maxvalue
            DiffLoadinterval(j) ← BaseLoadinterval(j) - Load(j)
        startpoint ← endpoint
        endpoint ← startpoint + interval
        iinterval ← iinterval + 1
    interval ← interval - 1
for j ∈ 1...maxinterval

```

$$\begin{aligned} Minfunc_j &\leftarrow \sum_k DiffLoad_j(k) + l * i_j \\ MinInt &\leftarrow \operatorname{argmin}(Minfunc) \\ MinCost &\leftarrow \min(Minfunc) \end{aligned}$$

Додатково введемо умову, що різниця між сусідніми репрезентативними значеннями не повинна бути меншою певного заданого порогу. Тобто для всіх інтервалів проводиться аналіз r_i , якщо $|r_i - r_{i+1}| < \varepsilon$, де ε – деякий поріг, тоді значення i та $(i+1)$ об'єднуються, а нове репрезентативне значення обирається як $\max(r_i, r_{i+1})$.

5.7. Розробка розкладів включення обладнання. Метод визначення моменту включення додаткового технічного засобу

В умовах необхідності заощаджувати витрати на утримання серверів, будь то орендна плата за використання ресурсів, які розташовані у хмарах, або енергетичні ресурси для забезпечення роботи власних серверів. Компанії оператору необхідно розробити план включення серверів, який би задовольняв всі потреби користувачів у поточний період часу.

Метод визначення моменту включення додаткового технічного засобу (серверу) полягає в тому, що із заданим інтервалом часу оцінюється поточна статистика, будується тренд, або лінійна апроксимація, зібраних даних про кількість заявок, що обслуговується, після чого робиться прогноз ймовірності того, що кількість заявок перевищить допустиму межу. Межею в даному випадку виступає допустима кількість заявок, яка може бути обслужена вже залученим увімкненим ресурсом.

За час для якого робиться оцінка добре вибрати час завантаження ще одного ресурсу, це дозволить, завчасно залучити необхідну кількість ресурсів.

Метод застосовується в процесі моніторингу системи, з метою виявлення моментів коли необхідно залучати додаткові ресурси.

Дана задача може вирішуватися як для системи в цілому, тоді оцінюється статистика кількості заявок на тарифікацію які надходять до

системи онлайн тарифікації, так і для підсистем, коли оцінюється статистика заявок-звернень до підсистем, системи тарифікації. Вибір точки застосування методу залежить від вибраної архітектури залучення серверів.

Отже, на основі поточної статистики навантаження, що створюється абонентськими заявками на тарифікацію; оцінки верхньої границі кількості заявок, які одночасно можуть оброблятися на потужностях доступних серверів буде визначена ймовірність того, що протягом заданого часу кількість заявок, що надійде не перевищить допустимого значення.

Алгоритм методу контролю достатності ресурсів системи для обробки заявок на тарифікацію

Вхідні дані:

3. інтервал $T1$ – інтервал часу, протягом якого буде проводитися аналіз статистики;
4. інтервал dt – інтервал дискретизації часу (малий інтервал);
5. дані системи моніторингу про кількість заявок на тарифікацію, які надійшли за малі інтервали часу, загалом $(T1/dt)$ результатів вимірів (Таблиця 5.1);

Таблиця 5.1

Кількість заявок на обслуговування

t	$t0-T1$	$t0-T1+T1/dt$	$t0-T1+2T1/dt$	$t0-T1+3T1/dt$...
X	$x(t)$	$x(t)$	$x(t)$	$x(t)$	

6. час протягом якого може бути завантажено сервера – T ;
7. таблиця значень Функції Лапласа (розподілу функції $N(0,1)$);
8. допустима ймовірність помилки, наприклад 10^{-4} ;
9. M – максимально допустима кількість заявок, які можуть обслуговуватися при поточній кількості ввімкнених серверів (ресурсів).

Включення серверу відбувається, якщо ймовірність недовостачі ресурсів вище допустимого значення Рис. 5.9

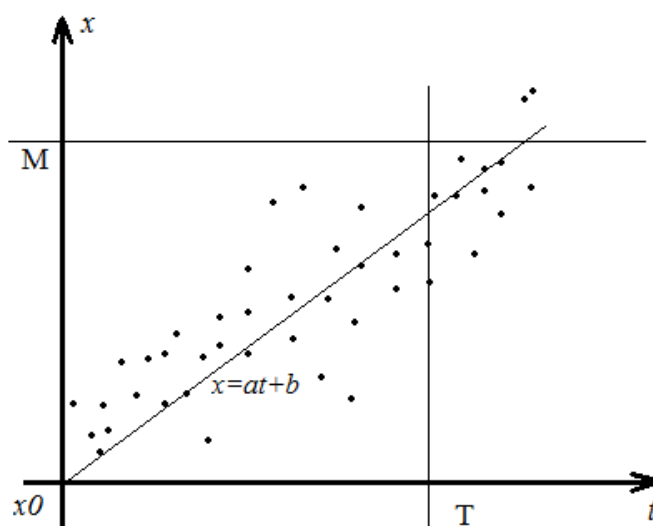


Рис. 5.9 Аналіз динаміки навантаження на сервер онлайн тарифікації.

Алгоритм методу включає в себе три основні кроки:

10. Аналіз статистичних даних за час (t_0-T) , де t_0 – поточний час, я якого виконується розрахунок. На основі статистичних даних системи моніторингу для пари значень (t, x) за методом найменших квадратів розраховується оцінка коефіцієнту a для прямої, яка апроксимує значення вхідного навантаження $x=at+b$

11. Оцінка ймовірності помилки протягом заданого часу розраховується за формулою (5.6).

$$P(\text{помилка за час } T) \leq 2P(N(0,1) \geq y) \quad (5.6)$$

де $y = ((M - x_0) - \hat{a} * T) / \sqrt{T}$,

x_0 – поточне навантаження на систему, M – допустиме навантаження на систему.

Ймовірність $P(N(0,1) \geq y)$, того що випадковий процес, який має нормальний розподіл, перевищить значення y знаходиться з таблиці значень Функції Лапласа.

12. Якщо ймовірність помилки протягом допустимого часу T ($P(\text{помилка за час } T)$) перевищує допустимий поріг, тоді відбувається включення додаткового серверу.

Застосування даного алгоритму дозволить контролювати динаміку зростання навантаження на сервер онлайн тарифікації та визначити момент коли необхідно залучити для обслуговування додаткові ресурси. Час T обирається виходячи з міркувань скільки потрібно часу від запуску до повної працездатності додаткового серверу.

Як видно з Рис. 5.9 лінія $x=at+b$ проходить нижче від точки з координатами (T, M) , оскільки даний метод враховує структуру випадкового процесу та середньоквадратичне відхилення яким характеризується випадковий процес надходження заявок на тарифікацію.

Дана функція перевірки є частиною системи моніторингу, запускається із заданою періодичністю для забезпечення надійності та безперебійної роботи системи, забезпечує контроль за достатністю ресурсів системи.

Оцінка ефективності

Спектр проблем, які можуть вирішуватися за допомогою запропонованого методу та алгоритмів досить широкий. Основною ознакою систем, для яких може бути застосований запропонований метод, є виконання великої кількості процедур, ініціатором яких є люди або інші програми. Виконання процедур здійснюється за допомогою програмного забезпеченні серверу, при цьому використовують технічні ресурси системи.

На сьогоднішній день проблема перевантаження вирішується за рахунок не допуску в систему надлишкової кількості запитів на тарифікацію, тобто якщо сервер є перевантаженим, то поступає сигнал на керуючий пристрій і тимчасово заявки на тарифікацію не приймаються. При цьому одночасно задіяні (знаходяться у режимі очікування) всі ресурси серверу, завантаження обладнання на 20-30% є нормальним в процесі обслуговування.

Був проведений експеримент роботи імітаційної моделі в двох режимах:

1. На обслуговування виділявся обмежений технічний ресурс, у разі перевантаження заявки відкидалися (Режим без застосування запропонованих методів).

2. Обслуговування заявок може проводитися від одного до трьох аналогічних серверів з обмеженими технічними ресурсами, на основі статистичних вибірки від оператора зв'язку був сформований розклад включення серверів відповідно до алгоритму пошуку моментів переключення, далі було згенеровано вхідний потік максимально наближений до реального. Включення серверів відбувалося за розкладом, а за методом включення додаткового технічного обладнання кожен три хвилини відбувалась перевірка достатності ресурсів для обслуговування вхідного потоку.

За результатами роботи імітаційної моделі було проведено дослідження використання електричної енергії, яке показало скорочення споживання енергії на 60% (Рис. 5.10), В той самий час кількість втрачених заявок скоротилася в 5 разів від п'яти відсотків до одного.

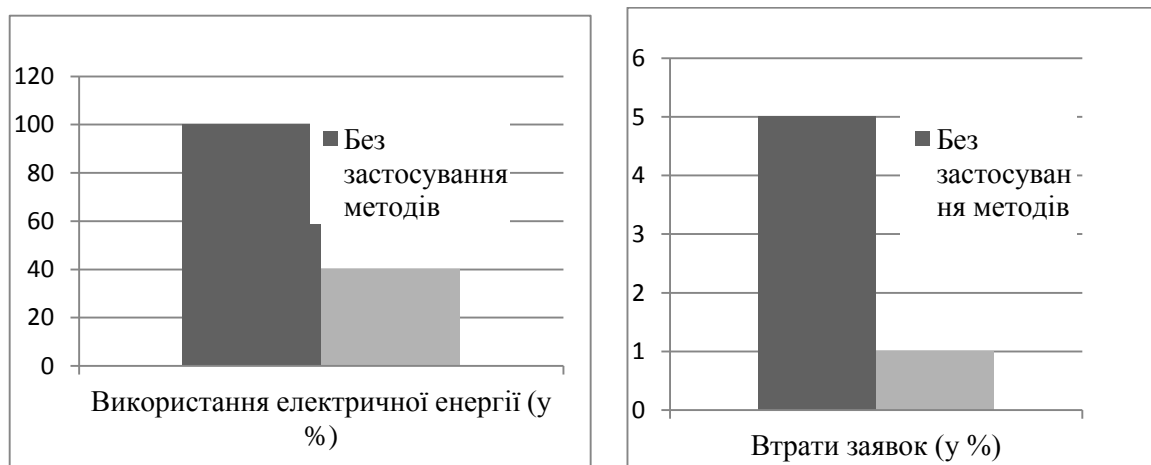


Рис. 5.10 а) Порівняння використання електричної енергії; б) Порівняння кількості втрачених заявок на тарифікацію

Розглянуто проблеми організації роботи системи онлайн тарифікації. Запропоновано підхід до системи керування роботою системи онлайн тарифікації, який полягає в тому, що система розглядається як мережа, керування потоками в мережі здійснюється із застосуванням системами балансування, де кількість доступних вузлів варіюється залежно від навантаження, яке потребує обслуговування (заявки на тарифікацію).

Запропоновано метод визначення моменту включення додаткового технічного засобу, який дозволяє проводити оцінку динаміки вхідного навантаження та поточного стану технічних засобів, та завчасно увімкнути додатковий ресурс, та запобігли перевантаженню існуючих ресурсів.

Проведене моделювання показало, що скорочується споживання електричної енергії на 60%, а втрати заявок через зайнятість серверу зменшуються у 5 разів від 5% до 1%.

5.7.1. **Механізми моніторингу платформи**

Завдяки можливості моніторингу різних складових елементів платформи, вона може адаптувати і масштабувати функції мережі, відповідно до різних метрик, де сутності прийняття рішень порівнюють отримані тригери з вимогами описів сервісів окремих елементів системи. Платформа надає можливість відділяти аспекти моніторингу функціонування мобільної мережі від тих, що відносяться до інфраструктури NFV. При цьому підтримуються існуючі механізми керування та оркестрування.

Інтерпретація подій моніторингу MCE і OSS/BSS з фізичної LEPC може використовуватись для створення vEPC в NFV інфраструктурі, а також для створення єдиного елемента або компонента vEPC. Доступні механізми моніторингу платформи можуть надаватись на різних рівнях:

1) *На рівні VNF*. Функції моніторингу вбудовуються за допомогою VNF і вся зібрана інформація передається в VNFM або EM. VNFs може направляти дані сутностям OSS/BSS і MCE разом з подіями моніторингу, які мають можливість приймати керуванські рішення для загальної мережі, враховуючі віртуалізовані функції.

2) *На рівні EM*. Рішення про події реалізуються через EM, з урахуванням даних з VNF, і направляються на VNFM або OSS/BSS і MCE для розповсюдження на всю мережу.

3) *На рівнях OSS/BSS і MCE.* Функція моніторингу не реалізується в ЕМ, або може перетинати кілька ЕМ. Вона може грати роль точки детектування або рішення. Події моніторингу LEPC також передаються на OSS/BSS, що дозволяє змінювати деякі параметри або елементи віртуалізованої мережі мобільного зв'язку.

4) *На рівні NFVO.* Якщо робота віртуалізованого сервісу складається з кількох пов'язаних між собою VNF із загальним кінцем, тобто зазнає впливу на рівні мережевого сервісу, це надає можливість спільно розглядати декілька звітів про моніторинг від окремих VNF, покращуючи NFVO. Тобто, можна впливати на функціональні потреби мережевих сервісів.

5) *На рівні VNFM.* Рішення про масштабування приймається на основі подій VNF, відповідно до параметрів моніторингу і масштабування. Даний об'єкт використовує інформацію про ці події і виконує необхідні дії по відношенню до інфраструктури NFV, отримуючи дані від менеджера VIM.

Моніторинг вищерозглянутої платформи, направлений на загальну поведінку масштабування віртуальних сервісів, може відбуватись трьома шляхами:

- Автоматичне масштабування – VNFM відслідковує події і виконує операцію масштабування при певних умовах;
- Масштабування за вимогою – VNF або ЕМ здійснює масштабування, відправляючи відповідний запит на VNFM;
- Ініціювання відправником (OSS/BSS або MCE) запиту керування до VNFM через NFVO.

Для моніторингу роботи площини керування і площини користувача, встановлюються точки спостереження S1-U і S5, які аналізують продуктивність трафіку даних та S1-MME і S10, які відслідковують продуктивність навантаження сигналізації та активність абонентів на користувацькому обладнанні за допомогою відповідних операцій. Опорні точки відправляють інформацію моніторингу на OSS/BSS.

5.8. Методологія забезпечення ефективного обслуговування гібридних телекомунікаційних сервісів у ГТС

Методологію забезпечення якості обслуговування, яка дозволяє ефективно керувати процесом обслуговування гібридних телекомунікаційних сервісів у гетерогенному телекомунікаційному середовищі

Відповідно до специфікації ITU-T M.3371(10/16) наразі розробляються методи моделі та підходи до керування системою обслуговування телекомунікаційних сервісів у багатошарному середовищі, виділяють три основні групи систем забезпечення: фізичні ресурси, хмарні ресурси та гетерогенна система керування інформаційними, обчислювальними та телекомунікаційними ресурсами. Основні складові процесу обслуговування гібридних телекомунікаційних сервісів показано на Рис. 5.11.



Рис. 5.11 Складові процесу обслуговування гібридних телекомунікаційних сервісів

В процесі керування обслуговуванням гібридних телекомунікаційних сервісів виникає потреба операціях з великою кількістю вхідних та вихідних даних. З метою систематизації було виділено три основні групи множин вхідних параметрів: група множин динамічно змінюваних параметрів, група множин граничних значень параметрів, які контролюються та використовуються у запропонованій сукупності методів:

Група динамічно змінюваних параметрів:

$V = \{V\}$ – множина вузлів доступних для обслуговування;

$E = \{e_{ij}\}$ – матриця зв'язності вузлів у досліджуваній момент часу;

$\Lambda = \{\lambda_{ij}\}$ – множина розрахункового навантаження j -го типу сервісу на i -й вузол обслуговування;

$R = \{R_{ij}\}$ – матриця доступних ресурсів, R_{ij} – кількість доступних ресурсів j -го типу в i -му вузлі обслуговування;

$N = \{n_i\}$ – множина кількостей паралельних каналів обслуговування у вузлі, n_i – кількість каналів для обслуговування у i -му вузлі;

$M = \{\mu_i\}$ – множина інтенсивностей обслуговування у вузлі, μ_i – поточна розрахована інтенсивність обслуговування заявок у i -му вузлі.

Група множин граничних значень параметрів:

$D = \{d_{ij}\}$ – множина граничних значень затримки у вузлі для різних типів сервісів, d_{ij} – граничне значення для затримки j -го типу сервісу в i -му вузлі обслуговування;

$P = \{P_i\}$ – множина граничних значень для показника втрат пакетів для різних типів сервісів, P_i – допустима ймовірність втрат пакетів i -го типу сервісу у вузлі обслуговування;

$S = \{(S_{min\ i}, S_{max\ i})\}$ – множина пар граничних значень коефіцієнту оптимального використання ресурсів i -го вузла, де $S_{min\ i}$ – мінімальне оптимальне значення коефіцієнту використання ресурсів (від 0 до 100%) i -го вузла, $S_{max\ i}$ – максимальне оптимальне значення коефіцієнту використання ресурсів (від 0 до 100%) i -го вузла.

Група рідко змінюваних параметрів:

$F = \{f_i(r, \rho)\}$ – множна функцій енергоспоживання вузлів, які залежать від навантаження та конфігурації вузла.

$r = \{r_{conf\ ij}\}$ – матриця можливих конфігурацій вузла, де i – номер вузла, j – номер конфігурації

T – множина допоміжних параметрів часу, які використовуються для формування статистичних вибірок та формування розкладу.

Група множин, що утворюється для організації розрахунків:

M_{max} - множина максимальної кількості заявок, які можуть бути обслужені у i -му вузлі за вибраної конфігурації системи, розраховується, розраховується для кожного типу сервісу на основі множин граничних значень (Рис. 5.12).

s – множина допустимих довжин черг для різних типів трафіків залежно від поточної конфігурації системи, розраховується на основі вимог до затримки, коефіцієнту використання ресурсів, та поточної інтенсивності обслуговування

l – множина граничних значень кількості заявок у черзі, для яких не застосовується метод раннього перевантаження системи

$v = \{v_{Rij}\}$ - множина об'єму ресурсу j -го типу, який використовується для обслуговування однієї заявки у i -му вузлі.

Запропонована методологія, базується на аналізі статистичних даних функціонування системи, дозволяє передбачати невідповідність діючої системи обслуговування вимогам потоків у наступний період часу, включає в себе методи керування як вхідним навантаженням так і ресурсами обслуговування, що відповідає можливостям віртуалізованого простору обслуговування.

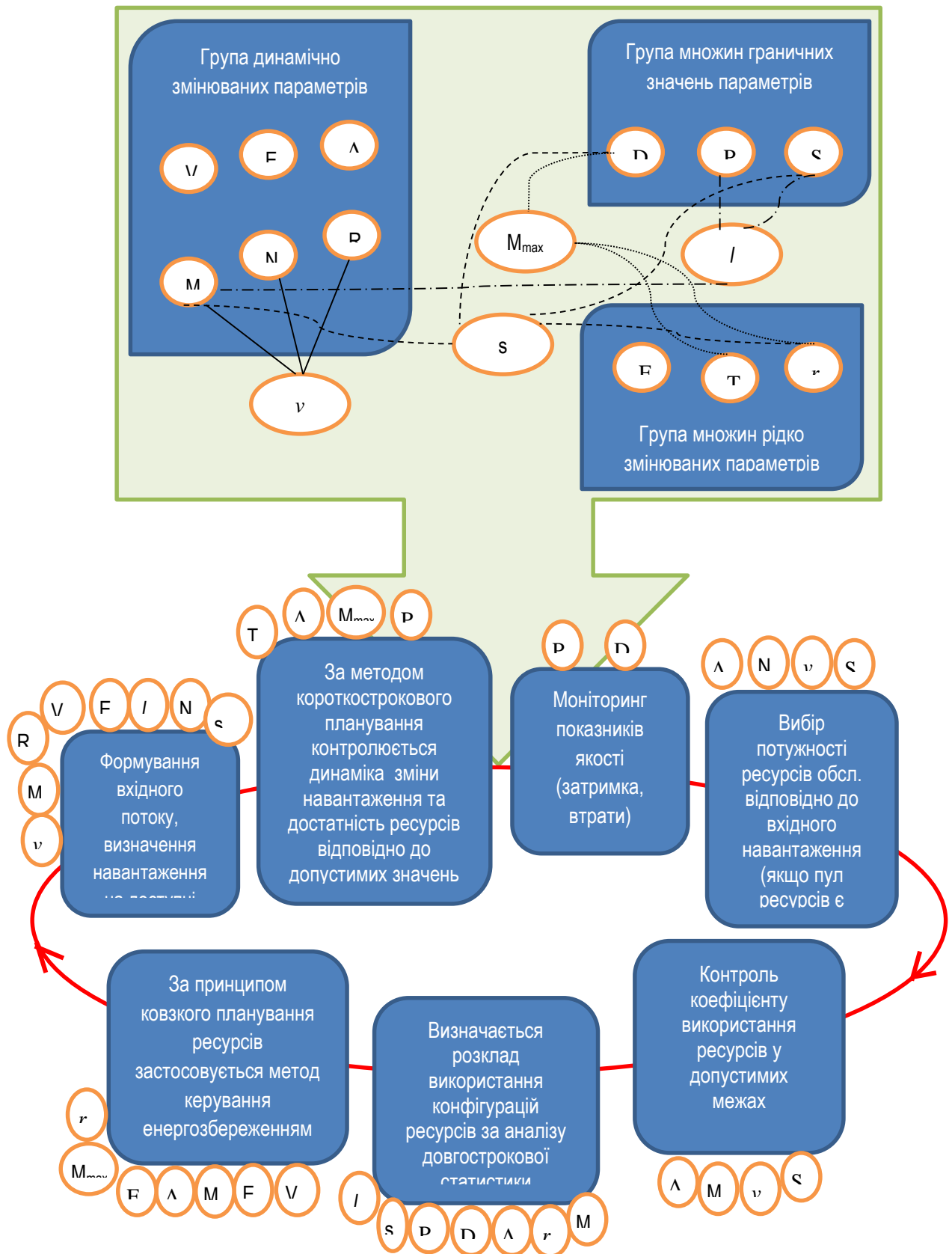


Рис. 5.12 Методологія забезпечення якості обслуговування гібридного телекомунікаційного сервісу

Висновки

Запропоновано вперше метод побудови розкладу залучення ресурсів на основі довгострокової статистики із забезпеченням показників якості обслуговування гібридних телекомунікаційних сервісів та ефективного використання ресурсів обслуговування, який за рахунок аналізу довгострокової статистики, ітераційної оцінки статистичних характеристик для інтервалів різної довжини, дозволяє сформувати розклад зміни конфігурації ресурсів обслуговування для вузла обслуговування у гетерогенному телекомунікаційному середовищі, та зменшити коефіцієнт простою ресурсів на 15%.

Вдосконалено авторегресійний метод із ковзковим математичним очікуванням основна, відмінність якого полягає у визначенні на основі довгострокової статистики інтервалу аналізу статистики, який забезпечує достовірність прогнозу із заданою ймовірністю, та мінімізує кількість точок аналізу, а відповідно й операцій при прогнозуванні у режимі реального часу, точність прогнозу, порівняно з фіксованим інтервалом аналізу, збільшилася на 3%, час розрахунку прогнозованого значення у режимі реального часу зменшився на 5%.

Запропоновано технологію керування гетерогенним телекомунікаційним середовищем, де обслуговування гібридних телекомунікаційних сервісів ведеться із застосуванням програмного забезпечення у багато хмарних дата центрах, технологія дозволила уникнути зниження якості обслуговування під час сплесків перевантаження та утримати показники якості обслуговування на заданому рівні за умови дотримання коефіцієнта використання ресурсів у заданих межах.

РОЗДІЛ 6

ІМІТАЦІЙНІ МОДЕЛІ ЗАПРОПОНОВАНИХ МЕТОДІВ

6.1. Дослідження ефективності методу оптимального вибору обчислювальних ресурсів для білінгових систем

Лояльність абонентів напряду залежить від рівня якості наданих послуг; пропозиції надання нових сервісів; швидкості роботи сервісів. Якість повинна, якщо не стабільно покращуватись, то утримуватись на достатньо високому рівні.

У сучасному світі об'єми інформації, що передається, невпинно збільшуються, що, у свою чергу, потребує постійного розширення пропускних можливостей мобільної мережі. Як наслідок, навантаження на внутрішні підсистеми обробки трафіку збільшуються, при цьому критично важливо уникнути перенавантажень на компонентах мережі, втрати даних і черг на обробку замовлень.

Дохід оператора мобільного зв'язку напряду залежить від організації процесу контролю фінансового стану рахунку (балансу) абонента - точності списання коштів абонента у режимі реального часу. Це дозволяє уникнути дебіторської заборгованості і підвищити комфорт користування послугами оператора.

Описані вище фактори визначають роль білінгової системи, як одного з головних компонентів в організації всього бізнес-процесу.

Якість послуг і дохід операторів мобільного зв'язку суттєво залежить від того, наскільки правильно розрахована продуктивність та ефективність білінгових систем, що обробляють потоки вхідних заявок на обслуговування викликів тарифікації. Для білінгових підсистем, які надають послуги по списанню коштів абонента у режимі реального часу, важливою характеристикою роботи є обробка вхідних заявок в найкоротші терміни, без створення черги.

Щоб на вхідних інтерфейсах білінгової підсистеми обслуговування викликів тарифікації не виникали черги оброблюваних даних, необхідно контролювати об'єм вхідного потоку, підтримувати оптимальне навантаження і своєчасно планувати розширення білінгових підсистем, що обробляють такий потік.

Вибраний метод дозволяє провести розрахунок рекомендованого значення інтенсивності вхідного потоку на існуючій білінговій підсистемі, порівняти його з доступними метриками продуктивності і спланувати рекомендоване розширення білінгових модулів при прогнозованому збільшенні вхідного потоку.

6.1.1. Метод формування вхідного потоку навантаження для ефективного використання ресурсів обслуговування

Основна ідея методу полягає в тому, щоб виходячи з ергодичного розподілу для можливих станів системи сформулювати вимоги до середнього значення вхідного навантаження, що дозволить максимально ефективно використовувати наявні фізичні ресурси обслуговування вхідного потоку заявок.

Процес обслуговування моделюється як n -канальний обслуговуючий пристрій, час обслуговування заявки у каналі є випадковою величиною розподіленою за законом Пуассона.

Вхідні дані.

n – кількість каналів для одночасного обслуговування заявок.

μ – інтенсивність обслуговування заявки,

G – кількість ресурсів залучених для обслуговування заявок,

v^g – об'єм g -го ресурсу необхідний для обслуговування у блоці однієї заявки, $g = \overline{1, G}$)

V^g – доступний об'єм ресурсу g -го ресурсу який спільно використовується заявками.

s – допустима кількість запитів у черзі на обслуговування.

R – відсоток заявок, які обслуговуються у системі не більше допустимого часу затримки, визначається експертами.

l – кількість запитів у черзі, до досягнення якої блокується надходження запитів до системи, відповідно до алгоритмів раннього попередження перевантажень.

Вихідні дані.

λ - рекомендоване значення для інтенсивності вхідного потоку, який буде направлено на обслуговування у s -канальний обслуговуючий пристрій.

Застосування запропонованого методу складається з двох етапів.

Етап 1. Для багатоканальної системи обслуговування, відповідно до моделей К. Жернового, необхідно знайти ергодичний розподіл кількості заявок у системі за формулами:

$$p_0 = \frac{1 - \beta^{s-l}}{A_n(\alpha, \beta)}, \quad \beta \neq 1, \quad \alpha = \lambda/\mu, \quad \beta = \lambda/n$$

$$A_n(\alpha, \beta) = (1 - \beta^{s-l}) \sum_{k=0}^n \frac{\alpha^k}{k!} + \frac{\alpha^n}{n!} \left(\frac{\beta - \beta^{s-l+1}}{1 - \beta} - (s-l)\beta^{s+1} \right)$$

$$p_k = \frac{\alpha^k}{k!} p_0 \quad (k = \overline{1, n})$$

$$p_{n+k} = \frac{\alpha^n}{n!} \beta^k p_0 \quad (k = \overline{1, l})$$

$$p_{n+k} = \frac{\alpha^n}{n!} \frac{\beta^k - \beta^s}{1 - \beta^{s-l}} p_0 \quad (k = \overline{l+1, s-1})$$

$$p_{n+s} = \frac{\alpha^n}{n!} \frac{(1 - \beta)\beta^s}{A_n(\alpha, \beta)} \quad (k = \overline{n+l+1, n+s-1})$$

Якщо $\beta = 1$, $\alpha = \lambda/\mu$, $\beta = \lambda/n$, тоді

$$p_k = \frac{n^k}{k!} p_0 \quad (k = \overline{1, n}); \quad p_{n+k} = \frac{n^n}{k!} p_0 \quad (k = \overline{1, l});$$

$$p_{n+s} = \frac{n^n}{n!} p_0 \quad (s-l-1)p_{n+s} \Rightarrow$$

$$\Rightarrow p_{n+s} = \frac{n^n}{n! (s-l)} p_0 \quad (k = \overline{n+l+1, n+s-1});$$

$$p_{n+k} = \frac{n^k}{n!} p_0 - (k-l)p_{n+s} = \frac{n^n}{n!} \frac{s-k}{s-l} p_0 \quad (k = \overline{l+1, s-1})$$

Етап 2. Розв'язання оптимізаційної задачі пошуку максимального вантаження, що забезпечить виконання умов на допустиму кількість ресурсів обслуговування.

$$\lambda \rightarrow \max$$

$$\begin{cases} 4 * \left(\sum_{i=1}^n i v_k^g p_i + n v_k^g \sum_{i=n+1}^s p_i \right) \leq V^g, g = \overline{1, G} \\ \sum_{i=1}^s p_i \leq R \end{cases}$$

У білінговій підсистемі балансування вхідного потоку відбувається з використанням модифікованої схеми циклічного розбору (Round Robin), для розрахунку рекомендованого значення інтенсивності вхідного потоку всієї системи буде достатньо проаналізувати метрики з одного DOCS сервера і масштабувати отримані значення на всю підсистему.

Для підстановки вхідних даних буде використана метрика DOCS-OCS-Congestion, зібрана на стороні одного DOCS сервера (Diameter-сервер). В табл. 1 приведено відповідність метрик і лічильників білінгової підсистеми вхідним даним обраної моделі. Всі значення були взяті на підсистемі обслуговування викликів тарифікації інтернет трафіку в момент найбільшої завантаженості ВНТА.

Таблиця 6.1

Відповідність метрик системи вхідним даним моделі

Параметр	Відповідність метрики	Одночасне значення
n	DOCS-OCS-Congestion	500
μ	DOCS-OCS-Congestion	ВНТА=403/сек
G	DOCS-OCS-Congestion	2048Mb
v_k^g	DOCS-OCS-	4Mb

	Congestion	
V^g	DOCS-OCS-Congestion	1612Mb
s	DOCS-OCS-Congestion	500/сек
R	Визначено експертом	90%
l	DOCS-OCS-Congestion	450

Розрахунок оптимального значення інтенсивності вхідного потоку для існуючої системи

Для розрахунку оптимального значення λ проведено аналіз доступних у білінговій підсистемі метрик для найбільш завантаженого дня в 2017 році (24.11.2017 «Чорна п'ятниця») [165]. Максимальне значення інтенсивності вхідного потоку було розраховано у години найбільшого навантаження, на момент, коли метрики сигналізували про відмови в обслуговуванні (REJECTS) для відомої кількості сесій. Під час аналізу, для кожного запису була розрахована інтенсивність обслуговування заявки (μ) і одночасне значення інтенсивності вхідного потоку (λ). Базуючись на отриманих даних, можна зробити висновок, що максимальному допустимому значенню інтенсивності вхідного потоку для єдиного OCS процесу, при якому не виникає деградації сервісу (відсутній REJECTS) відповідає $\lambda = 860$ при $\mu = 430$. Дані розрахунку оптимального значення λ і μ занесені в табл. 6.2.

Таблиця 6.2

Дані розрахунку оптимального значення λ і μ

DATE	TIME	GPRS1	GPRS2	REJECTS	μ	λ
24.11.2017	00:00	1211112	16102756	0	301	601.93
24.11.2017	01:00	977895	10609138	0	201	401.95
24.11.2017	02:00	962188	9984962	0	190	379.95
24.11.2017	03:00	971583	9719210	0	186	371.96
24.11.2017	04:00	973601	9570451	0	183	365.96
24.11.2017	05:00	1020312	10224994	0	195	389.95
24.11.2017	06:00	1157775	11722724	0	224	447.95
24.11.2017	07:00	1346571	14490450	0	275	549.93
24.11.2017	08:00	1749654	23858101	3	445	889.89
24.11.2017	09:00	1631822	21215782	0	397	793.90
24.11.2017	10:00	1693628	22199421	0	415	829.90
24.11.2017	11:00	1753250	22878271	0	428	855.89
24.11.2017	12:00	1817606	23379187	1	437	873.89
24.11.2017	13:00	1873052	23470265	2	440	879.89

24.11.2017	14:00	1907032	23036244	1	433	865.89
24.11.2017	15:00	1920535	22464995	1	431	861.89
24.11.2017	16:00	1945217	22005529	0	416	831.90
24.11.2017	17:00	1978970	22815321	0	430	859.89
24.11.2017	18:00	1986669	22961832	1	433	865.89
24.11.2017	19:00	1966866	22527160	0	425	849.89
24.11.2017	20:00	1958086	21745802	0	412	823.90
24.11.2017	21:00	1919809	19183564	0	366	731.91
24.11.2017	22:00	1831609	16819098	0	324	647.92
24.11.2017	23:00	1731484	14263415	0	278	555.93

Графічне відображення розрахованої інтенсивності обслуговування заявки (μ) і одночасного значення інтенсивності вхідного потоку (λ) для єдиного OCS процесу зображено на рис. 6.1.

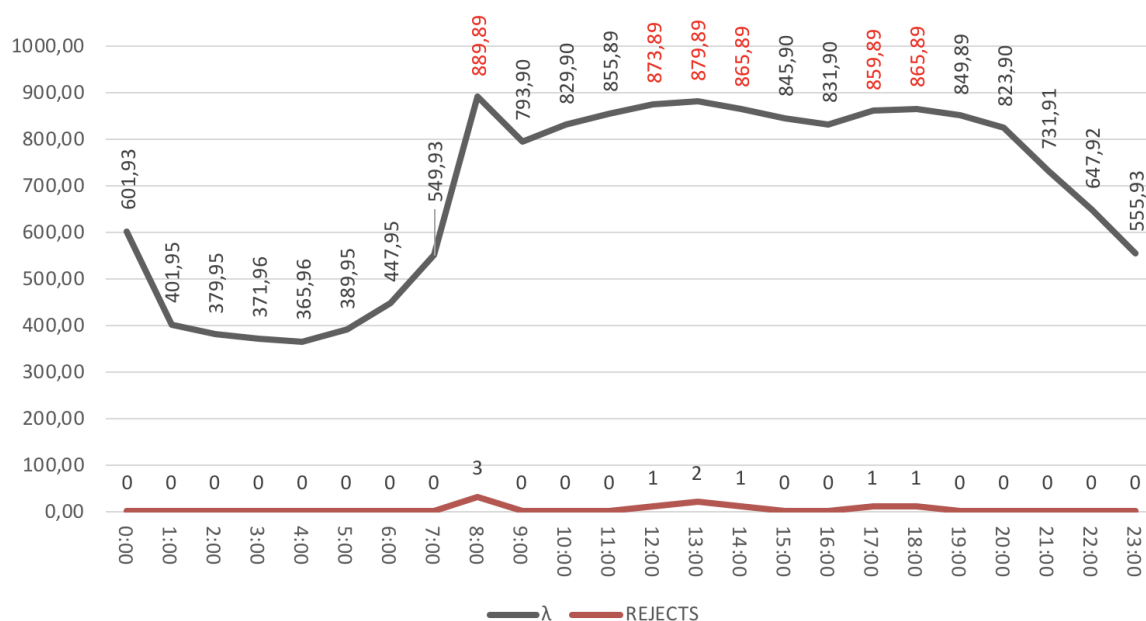


Рис. 6.1 Графік інтенсивності значення λ за 24.11.2017

Розрахунок оптимального значення інтенсивності вхідного потоку для системи, що масштабується

Відповідно до запропонованого розробником системи розширення фізичних компонентів, при масштабуванні системи – кількість OCS процесів, що виконуються на одному DOCS сервері, буде збільшено у 5 разів, при цьому конфігурація OCS процесу змінена не буде. Таким чином, для системи, що масштабується значення констант n , G , v_k^g , Vg , s , R , l , залишаться незмінними, а загальна кількість DOCS серверів буде збільшена.

Базуючись на розрахованому для такої системи значенні λ , можна розрахувати оптимальну кількість OCS процесів необхідних для підсистеми обслуговування викликів тарифікації інтернет трафіку, яка дозволить забезпечити доступність сервісів і цілісність даних, знизить вірогідність втрат і перенавантажень у системі після масштабування.

Враховуючи, що один OCS процес існуючої білінгової підсистеми до масштабування, при оптимальному значенні $\lambda = 860$ здатен обробляти трафік з інтенсивністю $\mu = 430$ заявок в секунду, всіх 16 DOCS серверів з одиничним запуском OCS процесом, оптимальна кількість трафіка, що обробляється буде дорівнювати 6880 сесій за секунду, що відповідає 24 768 000 сесіям в годину, опрацьованим всією підсистемою обслуговування викликів тарифікації інтернет трафіку.

Згідно з планами оператора мобільного зв'язку, інтенсивність вхідного потоку заявок повинна буде бути збільшеною до 50 000 000 сесій за годину. Виходячи з цього розробник білінгової системи дозволив розширити кількість OCS процесів до 30 штук зі збереженням конфігурації, що має забезпечити обробку зростаючого потоку вхідних заявок на білінгову систему. На рис. 6.2 зображено заплановане розширення білінгової підсистеми. Як видно, кількість DOCS серверів було зменшено до 6 штук, при цьому на кожному DOCS сервері запущено по п'ять OCS – процесів, які обробляють логіку дзвінка.

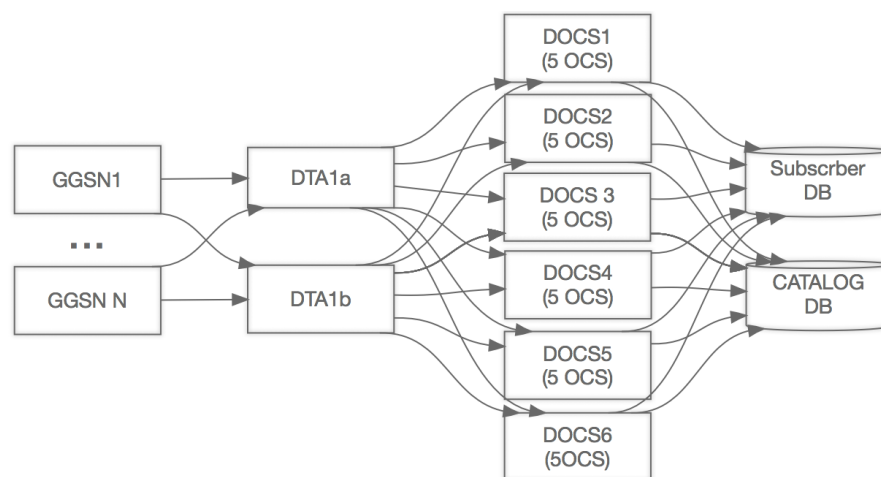


Рис. 6.2 Заплановане розширення білінгової підсистеми

Оскільки, значення інтенсивності вхідного потоку за годину і кількість паралельно працюючих OCS процесів у новій системі відомі, розраховуємо інтенсивність обробки трафіку за секунду, окремо на кожен процес. Використовуючи математичну модель, розраховуємо прогнозоване значення інтенсивності вхідного потоку λ . При 30 одночасно працюючих OCS процесах інтенсивність обробки трафіку за секунду буде $\mu = 463$. У табл. 6.3 відображено вхідні параметри для підставлення в імітаційну модель.

Таблиця 6.3

Вхідні параметри для підставлення в імітаційну модель

Параметр	Значення
n	500
μ	463/сек
G	2048Mb
v_k^g	4Mb
V^g	1612Mb
s	500/сек
R	90%
l	450

Підставивши ці дані в імітаційну модель, отримаємо $\lambda = 925.89$, що значно перевищує рекомендоване значення ($\lambda = 860$).

З вищеприведеного слідує, що запропонований варіант масштабування білінгової підсистеми буде недостатнім для задоволення вимог. Більше того, фізична поломка одного DOCS сервера спричинить багатократне збільшення трафіку на DOCS серверах, які продовжать роботу, що у свою чергу, обов'язково, відобразиться на кількості неопрацьованих замовлень і може призвести до повної відмови всієї білінгової підсистеми.

Керуючись принципом надмірності при архітектурному проектуванні розширення системи, додаємо два DOCS сервера. У такому випадку загальна кількість DOCS серверів, підключених до кластеру, буде дорівнювати 8, а кількість OCS процесів буде збільшено з 30 до 40 штук.

При наявності 40 OCS процесів, інтенсивність обробки трафіку за секунду для кожного OCS процесу буде $\mu = 347$. Після підставлення отриманого значення в імітаційну модель, маємо $\lambda = 693.91$, що задовольняє розраховане рекомендоване значення ($\lambda = 860$).

Навіть у випадку виходу з ладу одного DOCS сервера, прогнозоване значення інтенсивності вхідного потоку не перевищить рекомендоване значення і буде дорівнювати $\lambda = 793.90$, що перевищує відмовостійкість системи. На рис. 6.3 зображено покращене розширення з додаванням двох додаткових DOCS серверів.

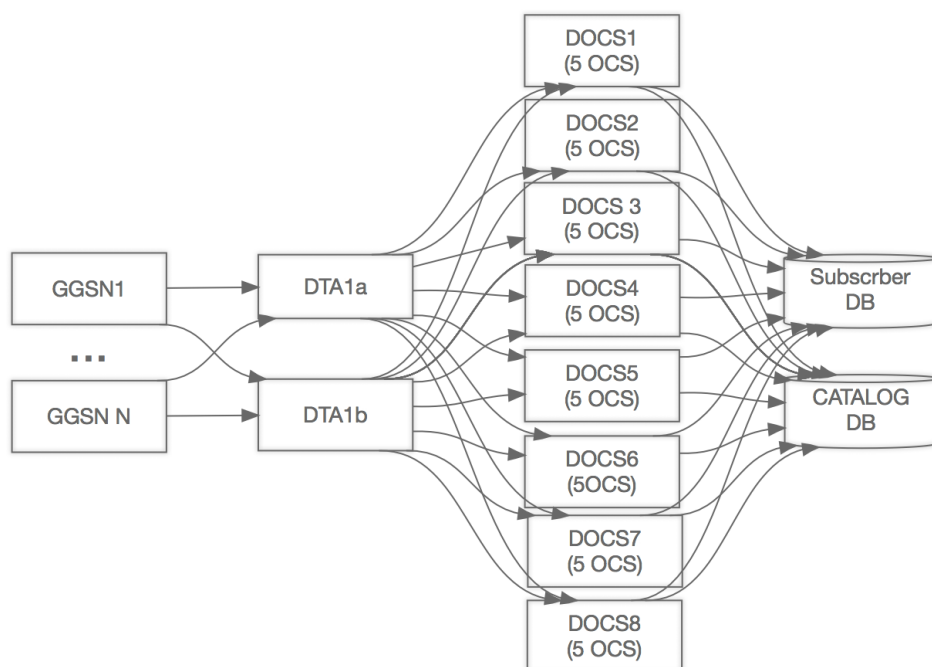


Рис. 6.3 Покращене розширення білінгової підсистеми

Розрахунок номінального значення інтенсивності вхідного потоку після масштабування білінгової системи

Під час проведення робіт по масштабуванню системи були враховані рекомендації по додаванню двох додаткових DOCS серверів. Білінгова. Підсистема була розширена до 8 DOCS серверів. Після переключення трафіку на нову підсистему, недоліків у роботі нової, розширеної підсистеми виявлено не було. Оператором мобільного зв'язку було прийнято рішення збільшити потік вхідних заявок до запланованої величини – орієнтовно, 50 000 000 заявок за годину. Система успішно витримала навантаження і

продовжує роботу у штатному режимі. Через кілька місяців, після збільшення потоку вхідних заявок було проведено порівняльний розрахунок інтенсивності вхідного потоку λ , для підтвердження дієздатності обраного методу. Отримані дані оптимального значення λ і μ , після розширення системи, занесені у табл. 6.4.

Таблиця 6.4

Дані порівняльного розрахунку λ і μ після розширення системи

DATE	TIME	GPRS1	GPRS2	REJECTS	μ	λ
11.04.2018	00:00	2725002	32205511	0	243	485.94
11.04.2018	01:00	2200264	21218275	0	163	325.96
11.04.2018	02:00	2164923	19969924	0	154	07.96
11.04.2018	03:00	2186062	19438419	0	150	299.96
11.04.2018	04:00	2190601	19140901	0	148	295.96
11.04.2018	05:00	2295701	20449988	0	158	315.96
11.04.2018	06:00	2604993	23445448	0	181	361.96
11.04.2018	07:00	3029785	28980900	0	222	443.95
11.04.2018	08:00	3936721	47716201	0	359	717.91
11.04.2018	09:00	3671600	42431565	0	320	639.92
11.04.2018	10:00	3810664	44398842	0	335	669.92
11.04.2018	11:00	3944812	45756542	0	345	689.92
11.04.2018	12:00	4089614	46758374	0	353	705.91
11.04.2018	13:00	4214366	46940530	0	355	709.91
11.04.2018	14:00	4290823	46072488	0	350	699.91
11.04.2018	15:00	4321204	44929989	0	342	683.92
11.04.2018	16:00	4376737	44011058	0	336	671.92
11.04.2018	17:00	4452683	44030643	0	337	673.92
11.04.2018	18:00	4470005	43737265	0	335	669.92
11.04.2018	19:00	4425450	42867923	0	328	655.92
11.04.2018	20:00	4405694	41305206	0	317	633.92
11.04.2018	21:00	4319571	38367128	0	296	591.93
11.04.2018	22:00	4121121	33638196	0	262	523.94
11.04.2018	23:00	3895840	28526830	0	225	449.95

Графічне відображення розрахованої інтенсивності обслуговування заявки (μ) і одночасного значення інтенсивності вхідного потоку (λ), для єдиного OCS процесу, після масштабування системи, зображено на рис. 6.4.

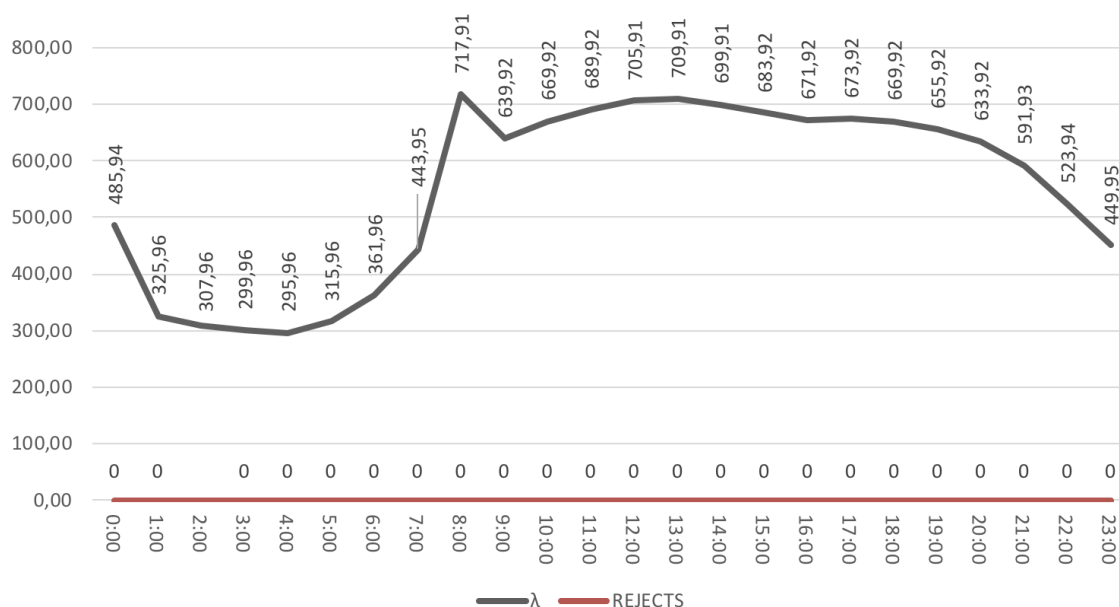


Рис. 6.4 Графік інтенсивності вхідного потоку (λ) за 11.04.2018

Після завершення усіх необхідних розрахунків бачимо, що інтенсивність вхідного потоку не перевищує розраховане рекомендоване значення $\lambda = 860$, відмови в обслуговуванні – цілком відсутні.

Для повноти аналізу проведемо розрахунок інтенсивності вхідного потоку так, якби рекомендоване додаткове розширення не було проведене, а загальна кількість DOCS серверів залишалась рівною шести.

Отримані дані прогнозованого значення λ і μ , після розширення системи, але з неповною кількістю DOCS серверів, занесені в табл. 6.5.

Графічне відображення прогнозованої інтенсивності обслуговування заявки (μ) и одночасного значення інтенсивності вхідного потоку (λ), для єдиного OCS процесу, після масштабування системи, але без запропонованого додаткового розширення зображено на рис. 6.5.

Таблиця 6.5

Данні порівняльного розрахунку λ і μ після розширення системи, при кількості DOCS серверів рівній 6

DATE	TIME	GPRS1	GPRS2	Передбач. REJECTS	μ	λ
11.04.2018	00:00	2725001,5	32205511	0	323	645.92
11.04.2018	01:00	2200264,3	21218275	0	217	433.95
11.04.2018	02:00	2164922,5	19969924	0	205	409.95
11.04.2018	03:00	2186061,8	19438419	0	200	399.95
11.04.2018	04:00	2190601,4	19140901	0	198	395.95
11.04.2018	05:00	2295701,2	20449988	0	211	421.95
11.04.2018	06:00	2604993,3	23445448	0	241	481.94
11.04.2018	07:00	3029785,2	28980900	0	296	591.93
11.04.2018	08:00	3936721,1	47716201	100	478	955.88
11.04.2018	09:00	3671600,4	42431565	100	427	853.89
11.04.2018	10:00	3810664	44398842	300	446	891.89
11.04.2018	11:00	3944812,3	45756542	300	460	919.89
11.04.2018	12:00	4089614,1	46758374	400	471	941.88
11.04.2018	13:00	4214366	46940530	450	474	947.88
11.04.2018	14:00	4290822,9	46072488	400	466	931.88
11.04.2018	15:00	4321203,9	44929989	300	456	911.89
11.04.2018	16:00	4376737,3	44011058	300	448	895.89
11.04.2018	17:00	4452683,3	44030643	300	449	897.89
11.04.2018	18:00	4470004,5	43737265	300	446	891.89
11.04.2018	19:00	4425449,6	42867923	200	438	875.89
11.04.2018	20:00	4405693,5	41305206	0	423	845.90
11.04.2018	21:00	4319571,1	38367128	0	395	789.90
11.04.2018	22:00	4121120,9	33638196	0	350	699.91
11.04.2018	23:00	3895840	28526830	0	300	599.93

Розрахунок прогнозу показує багатократне перевищення розрахованого рекомендованого значення ($\lambda = 860$). Також за допомогою методу екстраполяції були спрогнозовані відмови в обслуговуванні (REJECTS).

Після проведеного розширення білінгової підсистеми обслуговування викликів тарифікації, можна зробити висновок, що запропонований метод вибору обчислювальних ресурсів для обслуговування білінгових систем за умови коливання навантаження є оптимальним, початковий вибір цього методу був зроблений вірно.

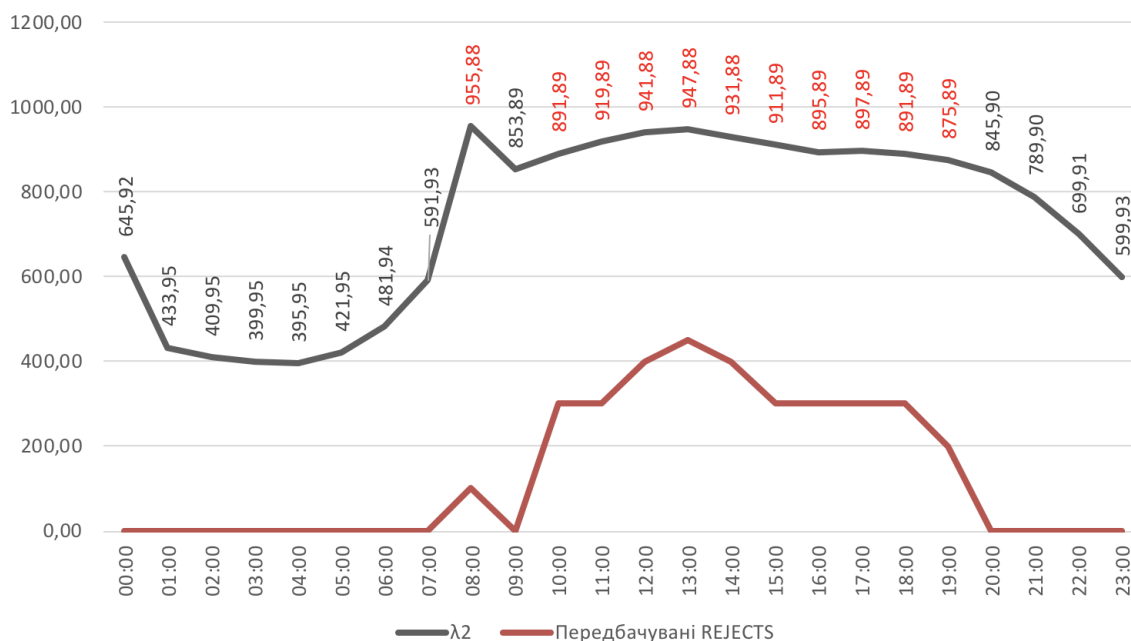


Рис. 6.5 Графік інтенсивності вхідного потоку (λ)

Запропонований метод і модель оптимального вибору обчислювальних ресурсів, для обслуговування білінгових систем за умови коливання навантаження є ефективними, у повній мірі враховують особливості архітектурних параметрів і можуть бути застосовані у дослідженні для масштабування системи і контролю навантаження білінгової системи.

У ході проведення дослідження масштабування білінгової системи було виявлено, що запропонований варіант масштабування білінгової системи не здатний у повній мірі обробляти запланований до збільшення вхідний потік заявок.

З огляду на виявлений у ході дослідження суттєвий недолік масштабування білінгової системи, було запропоновано і реалізовано додаткове збільшення ємності DOCS серверів, що дозволило забезпечити доступність сервісів і збереження цілісності даних, підвищило відмовостійкість і знизило вірогідність переванатажень системи.

6.2. Імітаційна модель MATLAB методу формування вхідного потоку навантаження для ефективного використання ресурсів обслуговування

Для розв'язання задачі розроблено 4 файли.

1. Файл `res.m` призначений для розрахунку імовірностей та визначення цільової функції `g`.

В завданні вигляд цільової функції, максимум якої слід знайти, не визначено. Обрано цільову функцію у вигляді суми всіх ймовірностей, починаючи з p_1 . Вибір можна змінити.

```
function [g]=res()
global n % кількість каналів для одночасного обслуговування
заявок
global s % допустима кількість запитів у черзі на
обслуговування
global l % кількість запитів у черзі, до досягнення якої
блокується надходження
global p % масив імовірностей
global alpha beta %

if beta==1
    A=alpha;
    for k=1:1:n
        A=A+alpha^k/factorial(k);
    end
    p(1)=1/A;
    for k=2:1:(n+1)
        p(k)=n^(k-1)/factorial(k-1)*p(1);
    end
    for k=1:1:1
        p(n+k+1)=n^(n)/factorial(k-1)*p(1);
    end
    for k=(n+1+1):1:(n+s-1)
        p(n+k+1)=n^(n)/factorial(n)*(s-1)*p(1);
    end
    for k=(1+1):1:(s-1)
        p(n+k+1)=n^(n)/factorial(n)*p(1)-(k-
1)*resm(n^(n)/factorial(n)*(s-1)*p(1));
    end
else
    A=alpha;
    for k=1:1:n
        A=A+alpha^k/factorial(k);
    end
    A=(1-beta^s)*A+alpha^n/factorial(n)*((beta-beta^(s-1+1))/(1-
beta)-(s-1)*beta^(s+1));
```

```

p(1)=(1-beta^(s-1))/A;
for k=2:1:(n+1)
    p(k)=alpha^(k-1)/factorial(k-1)*p(1);
    %p(k)=k;
end
for k=1:1:1
    p(n+k+1)=alpha^(n)/factorial(n)*p(1);
    %p(n+k+1)=k;
end
for k=(1+1):1:(s-1)
    p(n+k+1)=alpha^(n)/factorial(n)*(beta^k-beta^s)*p(1)/A;
    %p(n+k+1)=k;
end
for k=(n+1+1):1:(n+s-1)
    p(n+s+1)=alpha^(n)/factorial(n)*(1-beta)*beta^s/A;
    %p(n+s+1)=k;
end
g=sum(p);    % цільова функція
end

```

2. Файл `rosenbrock.m` призначений для надання всіх необхідних констант та ініціалізації робочих масивів. В ньому слід задати (записати в текстовому редакторі) наступні константи:

n – кількість каналів для одночасного обслуговування заявок.
 m_i – інтенсивність обслуговування заявки,
 G – кількість ресурсів залучених для обслуговування заявок,
 v_g – об'єм g -го ресурсу необхідний для обслуговування у блоці однієї заявки, $g = \overline{1, G}$

V_g – доступний об'єм ресурсу g -го ресурсу який спільно використовується заявками.

s – допустима кількість запитів у черзі на обслуговування.

R – відсоток заявок, які обслуговуються у системі не більше допустимого часу затримки, визначається експертами

l – кількість запитів у черзі, до досягнення якої блокується надходження запитів до системи відповідно до алгоритмів раннього упередження перевантажень.

```
function f = rosenbrock(x)
```

```

global n      % кількість каналів для одночасного обслуговування
заявок
global mu     % інтенсивність обслуговування заявки
global G      % кількість ресурсів залучених для обслуговування
заявок
global vg     % об'єм g-го ресурсу необхідний для обслуговування
у блоці однієї заявки
global Vg     % доступний об'єм ресурсу g-го ресурсу який спільно
використовується
global s      % допустима кількість запитів у черзі на
обслуговування
global R      % відсоток заявок, які обслуговуються у системі не
більше доп. часу
global l      % кількість запитів у черзі, до досягнення якої
блок.
global p      % масив імовірностей
global alpha beta %
global coeffs % масив вагових коефіцієнтів
global y1 y2
%x(1)=lambda - рекомендоване значення для інтенсивності вхідного
потoku

n=3;
y1=zeros(1,n);      % допоміжний масив
y2=zeros(1,s);      % допоміжний масив
coeffs=zeros(1,n); %формування масиву вагових коефіцієнтів
for i=1:1:n
    coeffs(1,i)=i;
end
mu=2;
G=3;
vg=[0.7 0.9 0.9];
Vg=[0.5 0.5 0.5];
s=4;
R=0.6;
l=3;
alpha=x(1)/mu;
beta=alpha/n;
f=res;

```

3. Файл unitdisk.m призначений для формування обмежень на кількість ресурсів обслуговування.

```

function [c, seq] = unitdisk(x)
global n      % кількість каналів для одночасного обслуговування
заявок
global G      % кількість ресурсів залучених для обслуговування
заявок
global vg     % об'єм g-го ресурсу необхідний для обслуговування
у блоці однієї заявки
global Vg     % доступний об'єм ресурсу g-го ресурсу який спільно
використовується

```

```

global s      % допустима кількість запитів у черзі на
обслуговування
global R      % відсоток заявок, які обслуговуються у системі не
більше доп.часу
global p      % масив імовірностей
global coeffs % масив вагових коефіцієнтів
global y1 y2

for i=1:1:n
    y1(1,i)=p(i+1);
end
for i=1:1:s
    y2(1,i)=p(i+1);
end
yy=y1*coeffs'; % скалярний добуток імовірностей на вагові
коефіцієнти
yy2=sum(y2)-R; % обмеження на суму імовірностей
c=yy*vg-Vg;    % обмеження на ресурси
c(1,n+1) = yy2; % загальний вектор обмежень
seq = [ ];     % обмеження на точну рівність відсутні

```

4. Файл `start.m` призначений для запуску програми на виконання. Використовується функція оптимізації `fmincon`. В перших квадратних дужках необхідно показати початкове значення λ .

```

options = optimset('Display','iter','Algorithm','active-set');
[x,fval] = fmincon(@rosenbrock,[0.1],...
    [],[],[],[],[],[],@unitdisk,options)

```

Для запуску програми слід переписати всі вказані файли в робочу папку Matlab і запустити файл `start.m`.

Результат розрахунку для вказаного прикладу:

```

>> start

```

			Max	Line search	Directional
First-order					
Iter	F-count	f(x)	constraint	steplength	derivative
optimality	Procedure				
0	2	10.3813	-0.03288		
1	5	5.52452	-0.01624	0.5	-99.9
75.7					
2	7	4.77695	-0.01031	1	-26.3
19.2					
3	10	4.28726	-0.005139	0.5	-19.2
8.25					


```

      4      12      3.90001    1.411e-005      1      -15.1
1.12
      5      14      3.90096    1.017e-010      1      12.2
0.00115
Local minimum found that satisfies the constraints.
Optimization completed because the objective function is non-
decreasing in
feasible directions, to within the default value of the function
tolerance,
and constraints were satisfied to within the default value of
the constraint tolerance.
<stopping criteria details>
Active inequalities (to within options.TolCon = 1e-006):
    lower      upper      ineqlin    ineqnonlin
                                     2
                                     3
x =      0.2853
fval =      3.9010

```

6.3. Імітаційна модель системи онлайн тарифікації з додатковим сервером в середовищі GPSS

В даному розділі запропонована імітаційна модель системи з додатковим сервером в середовищі GPSS, яка ілюструє процес обробки заявок на сервері мобільного оператора враховуючи особливості процесу тарифікації, а також дає інформацію про кількість оброблених та/або втрачених заявок на основному, додатковому сервері та час обслуговування заявок.

За результатами моделювання проаналізована доцільність використання системи обслуговування абонентів мобільного оператора з додатковим сервером.

Вхідні дані

Під час обробки вхідний потік заявок проходить п'ять функціональних блоків. При проходженні кожного з блоків, на обробку заявок виділяється певна кількість ресурсів – процесорного часу, оперативної пам'яті, постійної пам'яті, зайнятості каналу сигнальним трафіком.

Об'єм займаємих ресурсів для обробки заявок різних сервісів представлений у табл. 6.6.

Таблиця 6.6

Об'єм займаємих ресурсів для обробки заявок різних сервісів

Функціональний блок	Використання ресурсів			Час використання ресурсів, мс
FB1.1	S1	R1	1000	2
		R2	5	
		R3	1000	Depends on NW speed
		R4	0	
	S2	R1	1000	2
		R2	5	
		R3	1000	Depends on NW speed
	S3	R1	1000	2
		R2	5	
		R3	1000	Depends on NW speed
		R4	0	
	S4	R1	1000	2
		R2	5	
		R3	1000	Depends on NW speed
FB1.2	S1	R1	1000	2
		R2	1	
		R3	1000	Depends on NW speed
		R4	0	
	S2	R1	1000	2
		R2	1	
		R3	1000	Depends on NW speed
		R4	0	
	S3	R1	1000	2
		R2	1	
		R3	1000	Depends on NW speed
		R4	0	
	S4	R1	1000	2
		R2	1	
		R3	1000	Depends on NW speed
		R4	0	
FB2.1	S1	R1	0	
		R2	0	
		R3	0	
		R4	0	
	S2	R1	0	
		R2	5	
		R3	1000	
		R4	1000	300
	S3	R1	0	
		R2	6	
		R3	2000	
		R4	2000	3600
FB2.1	S1	R1	0	
		R2	0	

		R3	0	
		R4	0	
	S2	R1	0	
		R2	5	
		R3	1000	
		R4	1000	300
	S3	R1	0	
		R2	6	
		R3	2000	
		R4	2000	3600
	S4	R1	0	
		R2	3	
		R3	500	
		R4	500	1
FB2.2	S1	R1	0	
		R2	0	
		R3	0	
		R4	0	
	S2	R1	1000	300
		R2	5	
		R3	1000	
		R4	0	
	S3	R1	2000	3600
		R2	6	
		R3	2000	
		R4	0	
	S4	R1	500	1
		R2	3	
		R3	500	
		R4	0	
FB3	S1	R1	500	1000
		R2	100	
		R3	500	
		R4	0	
	S2	R1	1000	1000
		R2	200	
		R3	1000	
		R4	0	
	S3	R1	1000	1000
		R2	500	
		R3	1000	
		R4	0	
	S4	R1	600	1000
		R2	200	
		R3	600	
		R4	0	
FB5	S1	R1	200	200
		R2	100	
		R3	200	
		R4	0	

	S2	R1	1000	200
		R2	200	
		R3	1000	
		R4	0	
	S3	R1	2000	200
		R2	600	
		R3	2000	
		R4	0	
	S4	R1	400	200
		R2	150	
		R3	400	
		R4	0	

R1 – об’єм оперативної пам’яті, що використовується,

R2 – Процесорний час

R3 – Занятість каналу сигнальним трафіком

R4 – Об’єм постійної пам’яті

При обробці заявок SMS використовуються функціональні блоки FB1.1, FB1.2, FB2.1, FB2.2, FB3, FB5.

При збільшенні потоку заявок на сервер, ресурсів, що виділяються для обробки потоку може бути не достатньо. Тому заявки можуть втрачатися і якість обслуговування відповідно зменшується.

Якщо при проходженні наступного функціонального блоку кількість втрачених заявок перевищує пороговий рівень (імовірність відказу $P > X$), то потік заявок направляється на додатковий сервер.

Якщо заявка очікує своєї черги на обслуговування більше заданого часу, то вона також передається на додатковий сервер.

Для визначення кількості необроблених заявок, середній час обробки заявок, коефіцієнти використання каналів обслуговування скористаємось пакетом моделювання систем масового обслуговування GPSS.

Вхідний потік є Пуассонівським. Відомі середній час надходження та обробки заявок. Можливе одночасне надходження декількох заявок.

Середній час обробки заявки вибирається як $2\max(\tau)$, де τ – час найдовшого зайняття одного із ресурсів.

Необхідно визначити критерій якості обслуговування, тобто максимальну кількість заявок, які можуть бути не обслужені. При досягненні цього числа, наступні заявки, що надходять на основний сервер будуть передані на додатковий.

Так як, обчислювальні потужності не дозволяють провести моделювання для безкінечної кількості заявок, визначимо кількість заявок, що надходять на основний сервер.

Прийmemo кількість заявок, що надходять на основний сервер на протязі часу моделювання – 25000.

Прийmemo максимально допустимий час очікування заявки в черзі на обробку перед кожним функціональним блоком – 3 секунди, а на додатковому сервері – 5 секунд.

Лістинг 6.6 Програмний код в середовищі GPSS

```
*Параметри додаткового сервера
MEM_D STORAGE 100000;
DISK_D STORAGE 600000;
KANAL_D STORAGE 100000000000;
*Параметри основного сервера
MEM STORAGE 1000000;
DISK STORAGE 600000;
KANAL STORAGE 100000000000;
*Генерація заявок
GENERATE (POISSON(1,0.001))
*Заняття черги перед першим блоком
OCHERED01 QUEUE OCH01;
*Перевірка на наявність ресурсів
TEST LE 2000,R$MEM;
*вихід з черги
TEST LE M1,3,OUT;
DEPART OCH01;
ENTER KANAL;
ENTER MEM,2000;
ADVANCE 0.4;
LEAVE MEM,2000;
LEAVE KANAL;
* Заняття черги перед другим блоком
OCHERED02 QUEUE OCH02;
* Перевірка на наявність ресурсів
TEST LE 1500,R$MEM;
* вихід з черги
```

```

TEST LE M1,5,OUT;
DEPART OCH02;
ENTER KANAL;
ENTER MEM,2000;
ADVANCE 0.2;
LEAVE MEM,2000;
LEAVE KANAL;
* Заняття черги перед 3 блоком
OCHERED03 QUEUE OCH03;
* Перевірка на наявність ресурсів
TEST LE 1000,R$MEM;
* вихід з черги
TEST LE M1,5,OUT;
DEPART OCH03;
ENTER KANAL;
ENTER MEM,2000;
ADVANCE 0.1;
LEAVE MEM,2000;
LEAVE KANAL;
* Заняття черги перед 4м блоком
OCHERED04 QUEUE OCH04;
* Перевірка на наявність ресурсів
TEST LE 2000,R$MEM;
* вихід з черги
TEST LE M1,5,OUT;
DEPART OCH04;
ENTER KANAL;
ENTER MEM,2000;
ADVANCE 0.1;
LEAVE MEM,2000;
LEAVE KANAL;
* Заняття черги перед 5м блоком
OCHERED05 QUEUE OCH05;
* Перевірка на наявність ресурсів
TEST LE 2000,R$MEM;
* вихід з черги
TEST LE M1,5,OUT;
DEPART OCH05;
ENTER KANAL;
ENTER MEM,2000;
ADVANCE 0.3;
LEAVE MEM,2000;
LEAVE KANAL;
OUT QUEUE PEREH_OCHER;
TEST GE Q$PEREH_OCHER,50,OCHERED01;
SERVER2 ADVANCE 0.1;
QUEUE OCHERED01_SERV2;
* Перевірка на наявність ресурсів
TEST LE 2000,R$MEM_D;
* вихід з черги
TEST LE M1,5,VUX;
DEPART OCHERED01_SERV2;

```

```

ENTER KANAL_D;
ENTER MEM_D,2000;
ADVANCE 0.4;
LEAVE MEM_D,2000;
LEAVE KANAL_D;
QUEUE OCHERED02_SERV2;
* Перевірка на наявність ресурсів
TEST LE 2000,R$MEM_D;
* вихід з черги
TEST LE M1,5,VUX;
DEPART OCHERED02_SERV2;
ENTER KANAL_D;
ENTER MEM_D,2000;
ADVANCE 0.4;
LEAVE MEM_D,2000;
LEAVE KANAL_D;
QUEUE OCHERED02_SERV2;
* Перевірка на наявність ресурсів
TEST LE 2000,R$MEM_D;
* вихід з черги
TEST LE M1,5,VUX;
DEPART OCHERED02_SERV2;
ENTER KANAL_D;
ENTER MEM_D,2000;
ADVANCE 0.4;
LEAVE MEM_D,2000;
LEAVE KANAL_D;
QUEUE OCHERED03_SERV2;
* Перевірка на наявність ресурсів
TEST LE 2000,R$MEM_D;
* вихід з черги
TEST LE M1,5,VUX;
DEPART OCHERED03_SERV2;
ENTER KANAL_D;
ENTER MEM_D,2000;
ADVANCE 0.4;
LEAVE MEM_D,2000;
LEAVE KANAL_D;
QUEUE OCHERED04_SERV2;
* Перевірка на наявність ресурсів
TEST LE 2000,R$MEM_D;
* вихід з черги
TEST LE M1,5,VUX;
DEPART OCHERED04_SERV2;
ENTER KANAL_D;
ENTER MEM_D,2000;
ADVANCE 0.4;
LEAVE MEM_D,2000;
LEAVE KANAL_D;
QUEUE OCHERED05_SERV2;
* Перевірка на наявність ресурсів
TEST LE 2000,R$MEM_D;

```

```

* вихід з черги
TEST LE M1,5,VUX;
DEPART OCHERED05_SERV2;
ENTER KANAL_D;
ENTER MEM_D,2000;
ADVANCE 0.4;
LEAVE MEM_D,2000;
LEAVE KANAL_D;
VUX TERMINATE
GENERATE 30
TERMINATE 1

```

Аналіз отриманих результатів

Було згенеровано 22486 заявок за законом Пуассона з параметрами (1, 0,001) протягом 30 секунд машинного часу моделювання (рис. 6.6).

LABEL	LOC	BLOCK TYPE	ENTRY COUNT	CURRENT COUNT	RETRY
	1	GENERATE	22486	0	0

Рис. 6.6 Кількість згенерованих заявок

За результатами моделювання втрати на функціональних блоках на основному сервері склали: на першому функціональному блоці – 3250 заявок, на другому – 54, на третьому – 281, на четвертому – 112, на п'ятому 53.

Отже, загальна кількість втрачених заявок при обробці на основному сервері складає $3250+54+281+112+53 = 3750$ заявок, що складає 16% усіх заявок.

До першого функціонального блоку додаткового сервера надійшло 10959 заявок, тобто 48% усіх заявок (рис. 6.7).

FBLOK1	54	ENTER	10959	0	0
	55	ENTER	10959	0	0
	56	ADVANCE	10959	49	0

Рис. 6.7 Кількість заявок, що надійшла на додатковий сервер

За результатами моделювання втрати на функціональних блоках на додатковому сервері склали:

На першому функціональному блоці – 49, на другому – 0, на третьому – 25, на четвертому – 426, на п'ятому – 0.

Отже, загальна кількість втрачених заявок на додатковому сервері складає

$$49 + 0 + 25 + 426 + 0 = 500,$$

що становить 4,56% від усіх отриманих на додатковому сервері заявок.

Загальний час обробки усіх заявок на додатковому сервері складає сумму часу на обробку в кожному функціональному блоці та час на передачу заявок від основного до додаткового сервера (7 стовпець на рис. 6.8).

$$t_{\text{заг}} = 5,275 + 1,633 + 2,508 + 3,358 + 11,693 + 0,2 = 24,667 \text{ секунд.}$$

Тоді, в середньому на обробку однієї заявки витрачається 0,0022 секунди що є прийнятним результатом.

QUEUE	MAX	CONT.	ENTRY	ENTRY(0)	AVE.CONT.	AVE.TIME	AVE. (-0)
OCH01	8702	7837	22535	783	3505.090	4.666	4.834
OCH02	501	500	14698	14198	314.697	0.642	18.882
OCH03	699	698	14198	13500	288.980	0.611	12.420
OCH04	432	431	13500	13069	30.397	0.068	2.116
OCH05	1236	1235	12818	11583	582.417	1.363	14.148
PEREH_OCHER	20209	20209	20209	0	9120.827	13.540	13.54
OCHERED01_SERV2	9250	9201	20160	1166	3544.783	5.275	5.599
OCHERED02_SERV2	1461	1460	10910	9450	593.987	1.633	12.205
OCHERED03_SERV2	1500	1499	9450	7951	789.997	2.508	15.810
OCHERED04_SERV2	1971	1970	7926	5956	887.203	3.358	13.511
OCHERED05_SERV2	3944	3944	5530	1586	2155.357	11.693	16.395

Рис. 6.8 Час обробки заявок

Отже, система з додатковим сервером дає низькі показники втрат заявок, а саме 4,56% втрачених заявок, натомість втрати на основному сервері складають близько 16%. Час обробки однієї заявки, враховуючи затримку при передачі на додатковий сервер складає 0,0022с, що є задовільним показником.

6.4. Моделювання задачі розподілу ресурсів між гібридними телекомунікаційними сервісами сервісами

Для розв'язання задачі розподілу ресурсів на сервері оператора мобільного зв'язку необхідно мати такі дані: кількість прибутку, який може бути отриманий від обслуговування однієї заявки кожного типу, загальний об'єм ресурсів, яким розполагає сервер; кількість ресурсів, необхідних усім функціональним блокам для обслуговування заявок кожного типу; статистичні дані про надходження викликів на сервер у різні періоди часу.

Вирішимо задачу розподілу ресурсів між гібридними ТК сервісами для системи онлайн тарифікації, де обслуговування чотирьох типів заявок (sms, дзвінки, передача даних через мережу Інтернет, mms) проводиться у п'яти функціональних блоках. Розв'яжемо задачу для двох типів ресурсів: оперативної (R1) і постійної (R2) пам'яті сервера. Для прикладу візьмемо дані про кількість ресурсів, необхідних кожному функціональному блоку для обслуговування заявки певного типу, представлені в таблиці 6.7.

Таблиця 6.7

Розподіл ресурсів між функціональними блоками

№ ФБ			FB1.1	FB1.2	FB2.1	FB2.2	FB3	FB4.1	FB4.2	FB5
Resource usage	R1 (байти)	S1	1000	1000	0	0	500	1000	1000	200
		S2	1000	1000	0	1000	1000	1000	100	1000
		S3	1000	1000	0	2000	1000	1000	100	2000
		S4	1000	1000	0	500	600	1000	100	400
	R2 (байти)	S1	0	0	0	0	0	100	1000	0
		S2	0	0	1000	0	0	100	1000	0
		S3	0	0	2000	0	0	100	1000	0
		S4	0	0	500	0	0	100	1000	0

S1 – дзвінки;

S2 – sms-повідомлення;

S3 – mms-повідомлення;

S4 – Інтернет-контент.

В таблиці 6.3 представлені основні задачі, які виконують функціональні блоки. У першому, другому і четвертому функціональних блоках заявки можуть обслуговуватись двома способами. При цьому кількість ресурсів залежатиме від того, який зі способів був обраний.

Таблиця 6.7

Призначення функціональних блоків

Функціональний блок	Задачі
FB1.1	Gathering subscriber data from Permanent Storage
FB1.2	Gathering subscriber data from cache
FB2.1	State write/read to permanent storage
FB2.2	State write/read to cache
FB3	Rating
FB4.1	Online notification
FB4.2	Offline notification
FB5	Final debit

У таблиці 6.4 представлена кількість прибутку (коефіцієнт корисності для оператора), яку може отримати оператор від обслуговування однієї заявки кожного типу сервісу в умовних одиницях.

Таблиця 6.8

Прибуток від кожного сервісу

Сервіс	Величина прибутку
Sms	5 у.о.
Voice	2 у.о.
Data transfer	6 у.о.
mms	3 у.о.

Прибуток від кожного сервісу залежить від тарифів у даний момент часу, які не є фіксованими, тому дані в таблиці взяті для прикладу.

Для того, щоб розв'язати задачу розподілу ресурсів на сервері оператора мобільного зв'язку, необхідно ввести додаткові обмеження для кількості заявок кожного типу. Для цього потрібно аналізувати статистичні дані про кількість заявок в кожний період часу.

Для прикладу розглянемо співвідношення між кількістю заявок різних типів у період із 12 до 13 години (рис.6.9)

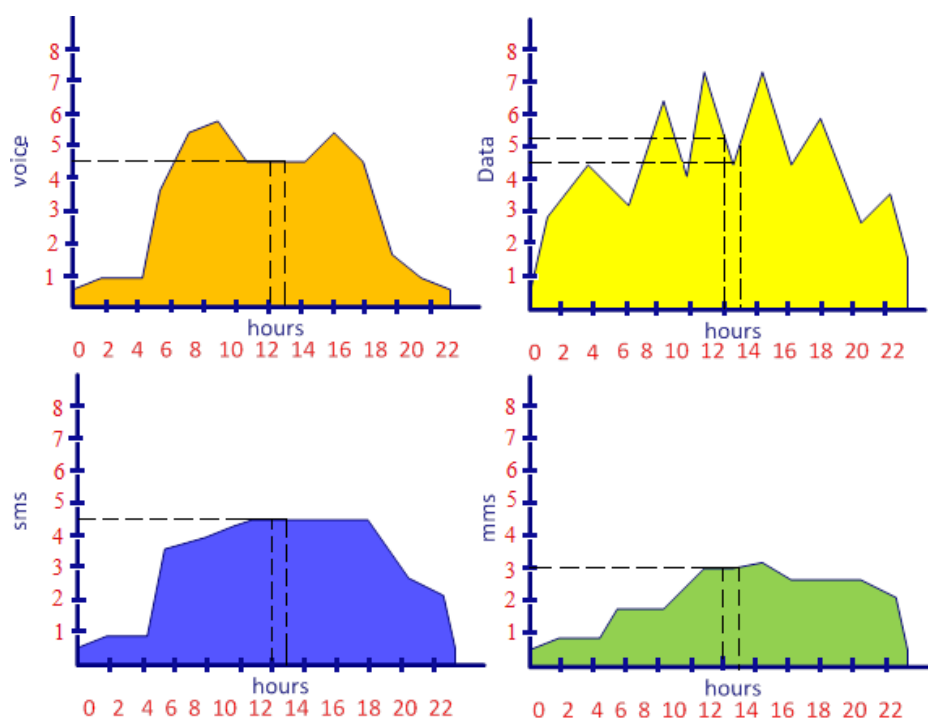


Рис. 6.9 Графіки навантаження по кожному сервісу

З рисунка видно, що кількість дзвінків, sms, mms та Інтернет-контенту у цей період складає в середньому 4,5; 4,5; 3 і 4,8 одиниць відповідно. Для того, щоб отримати обмеження на кількість заявок кожного типу, знайдемо співвідношення між ними у даний період часу.

Отримаємо такі співвідношення:

$$\frac{S1}{S2} = \frac{4,5}{4,5} = 1$$

$$\frac{S1}{S3} = \frac{4,5}{3} = 1,5$$

$$\frac{S1}{S4} = \frac{4,5}{4,8} = 0,94$$

$$\frac{S2}{S3} = \frac{4,5}{3} = 1,5$$

$$\frac{S2}{S4} = \frac{4,5}{4,8} = 0,94$$

$$\frac{S3}{S4} = \frac{3}{4,8} = 0,625$$

Розширивши діапазон значень від -0,5 до 0,5, отримаємо такі додаткові обмеження: $0,5 \leq \frac{S1}{S2} \leq 1,5$

$$1 \leq \frac{S1}{S3} \leq 2$$

$$0,44 \leq \frac{S1}{S4} \leq 1,44$$

$$1 \leq \frac{S2}{S3} \leq 2$$

$$0,44 \leq \frac{S2}{S4} \leq 1,44$$

$$0,125 \leq \frac{S3}{S4} \leq 1,125$$

Перейдемо до вирішення задачі розподілу ресурсів між сервісами з урахуванням економічної ефективності обслуговування заявок.

Математичне описання задачі наступне:

$$\sum_{i=1}^4 k_i S_i \rightarrow \max$$

$$\left\{ \begin{array}{l} \sum_{i=1}^4 k_i \left(\sum_{j=1}^8 v_{ij}^{R1} \right) \leq V_{R1} \\ \sum_{i=1}^4 k_i \left(\sum_{j=1}^8 v_{ij}^{R2} \right) \leq V_{R2} \\ 0,5 \leq \frac{k_1}{k_2} \leq 1,5 \\ 1 \leq \frac{k_1}{k_3} \leq 2 \\ 0,44 \leq \frac{k_1}{k_4} \leq 1,44 \\ 1 \leq \frac{k_2}{k_3} \leq 2 \\ 0,44 \leq \frac{k_2}{k_4} \leq 1,44 \\ 0,125 \leq \frac{k_3}{k_4} \leq 1,125 \end{array} \right.$$

Де k_1 - кількість заявок 1-го типу (дзвінки);

k_2 - кількість заявок 2-го типу (sms-повідомлення);

k_3 - кількість заявок 3-го типу (mms-повідомлення);

k_4 - кількість заявок 4-го типу (Інтернет-контент);

S_i - кількість прибутку, отриманого від обслуговування однієї заявки певного типу.

З таблиці видно, що є два варіанти проходження заявки через функціональні блоки, а саме через 1-й, 2-й та 4-й. Тому при вирішенні поставленої задачі будемо розглядати обидва способи.

Щоб визначити величини V_{R1} і V_{R2} (приблизну кількість оперативної і постійної пам'яті, яку може надати сервер), припустимо, що на сервер надходить по 10^6 заявок кожного типу. Скориставшись відомою інформацією про те, яку кількість ресурсів потребує кожен тип заявок у кожному функціональному блоці, підрахуємо загальну кількість оперативної і постійної пам'яті, необхідної для обслуговування усіх заявок.

Розглянемо випадок, коли заявка проходить наступний шлях:

$$FB1.1 \rightarrow FB2.1 \rightarrow FB3 \rightarrow FB4.1 \rightarrow FB5$$

1) Загальна кількість оперативної пам'яті

Кількість оперативної пам'яті, необхідної для обслуговування 1 заявки кожного типу.

$$S1: V_{R1}^1 = 1000 + 0 + 500 + 1000 + 200 = 2700$$

$$S2: V_{R1}^2 = 1000 + 0 + 1000 + 1000 + 1000 = 4000$$

$$S3: V_{R1}^3 = 1000 + 0 + 1000 + 1000 + 2000 = 5000$$

$$S4: V_{R1}^4 = 1000 + 0 + 600 + 1000 + 400 = 3000$$

$$V_{R1} = 10^6 \cdot (V_{R1}^1 + V_{R1}^2 + V_{R1}^3 + V_{R1}^4) = 10^6 \cdot (2700 + 4000 + 5000 + 3000) = 1,47 \cdot 10^{10}$$

2) Загальна кількість постійної пам'яті

Кількість постійної пам'яті, необхідної для обслуговування 1 заявки кожного типу.

$$S1: V_{R2}^1 = 0 + 0 + 0 + 100 + 0 = 100$$

$$S2: V_{R2}^2 = 0 + 1000 + 0 + 100 + 0 = 1100$$

$$S3: V_{R2}^3 = 0 + 2000 + 0 + 100 + 0 = 2100$$

$$S4: V_{R2}^4 = 0 + 500 + 0 + 100 + 0 = 600$$

$$V_{R2} = 10^6 \cdot (V_{R2}^1 + V_{R2}^2 + V_{R2}^3 + V_{R2}^4) = 10^6 \cdot (100 + 1100 + 2100 + 600) = 3,9 \cdot 10^9$$

Тобто загальна кількість оперативної і постійної пам'яті, яку можуть використовувати функціональні блоки для обслуговування заявок не може перевищувати ці числа.

Для вирішення задачі оптимізації цільової функції використовується пакет Microsoft Office (Excel). Щоб знайти параметри k_i ($i=1,4$), при яких цільова функція буде максимальною, скористаємося інструментом «Пошук рішення» на панелі інструментів Excel (рис.6.10).

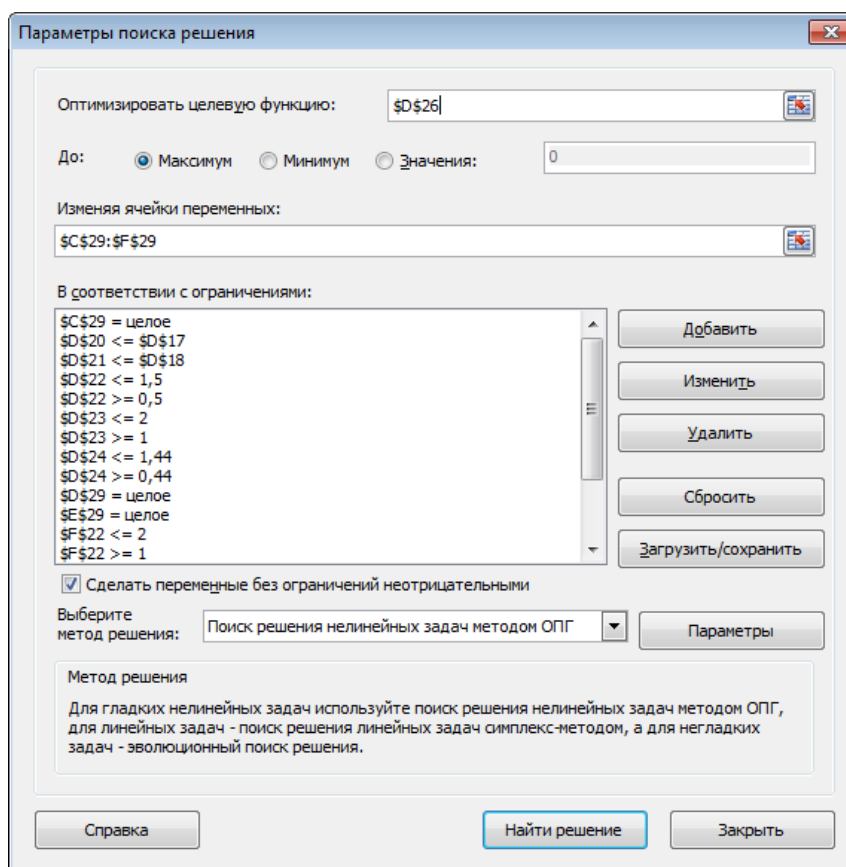


Рис. 6.10 Панель інструментів «Пошук рішення» MS Excel

Задаємо номер комірки з цільовою функцією, яку необхідно оптимізувати до максимуму і задаємо визначені вище обмеження. Крім того, необхідно вказати, що знайдені значення кількості заявок повинні бути цілими числами. Після визначення усіх необхідних умов знаходимо рішення задачі.

В результаті отримаємо такі значення кількості заявок:

Таблиця 6.10

Кількість заявок кожного типу

k1	k2	k3	k4
1247676	831784	831784	1281739

При цьому значення цільової функції, тобто максимального прибутку від обслуговування усіх заявок буде наступним:

$$\sum_{i=1}^4 k_i S_i = 16737869(\text{y. o.})$$

Перевіримо виконання усіх необхідних умов.

- 1) $\sum_{i=1}^4 k_i (\sum_{j=1}^8 v_{ij}^{R1}) = 1247676 \cdot (1000 + 0 + 500 + 1000 + 200) + 831784 \cdot (0 + 1000 + 0 + 100 + 0) + 831784 \cdot (1000 + 0 + 1000 + 1000 + 2000) + 1281739 \cdot (1000 + 0 + 600 + 1000 + 400) = 1,47 \cdot 10^{10} \leq V_{R1};$
- 2) $\sum_{i=1}^4 k_i (\sum_{j=1}^8 v_{ij}^{R2}) = 1247676 \cdot (0 + 0 + 0 + 100 + 0) + 831784 \cdot (1000 + 0 + 1000 + 1000 + 1000) + 831784 \cdot (0 + 2000 + 0 + 100 + 0) + 1281739 \cdot (0 + 500 + 0 + 100 + 0) = 3,56 \cdot 10^9 \leq V_{R2};$
- 3) $\frac{k_1}{k_2} = \frac{1247676}{831784} = 1,5 \quad 0,5 \leq 1,5 \leq 1,5;$
- 4) $\frac{k_1}{k_3} = \frac{1247676}{831784} = 1,5 \quad 1 \leq 1,5 \leq 2;$
- 5) $\frac{k_1}{k_4} = \frac{1247676}{1281739} = 0,973424 \quad 0,44 \leq 0,973424 \leq 1,44;$
- 6) $\frac{k_2}{k_3} = \frac{831784}{831784} = 1 \quad 1 \leq 1 \leq 2;$
- 7) $\frac{k_2}{k_4} = \frac{831784}{1281739} = 0,64895 \quad 0,44 \leq 0,64895 \leq 1,44;$
- 8) $\frac{k_3}{k_4} = \frac{831784}{1281739} = 0,64895 \quad 0,125 \leq 0,64895 \leq 1,125;$

Тобто усі необхідні умови виконуються при даних значеннях k_i .

Визначимо необхідний об'єм ресурсів для знайденої кількості заявок кожного типу за формулами:

$V_i^{R1} = k_i \sum_j v_{ij}^{R1}$ – кількість оперативної пам'яті;

$V_i^{R2} = k_i \sum_j v_{ij}^{R2}$ – кількість постійної пам'яті.

Отримаємо такі результати:

Таблиця 6.11

Розподіл ресурсів між заявками

R1	S1	S2	S3	S4
	3,37E + 09	3,33E + 09	4,16E + 09	3,8E + 09

R2	S1	S2	S3	S4
	1,25E + 08	9,15E + 08	1,75E + 09	7,69E + 08

Тобто кількість оперативної пам'яті для кожного типу сервісу дорівнює:

$$V_1^{R1} = 3,37 \cdot 10^9 \text{ - для дзвінків;}$$

$$V_2^{R1} = 3,33 \cdot 10^9 \text{ - для sms-повідомлень;}$$

$$V_3^{R1} = 4,16 \cdot 10^9 \text{ - для mms-повідомлень;}$$

$$V_4^{R1} = 3,8 \cdot 10^9 \text{ - для Інтернет-контенту.}$$

Кількість постійної пам'яті для кожного типу сервісу дорівнює:

$$V_1^{R2} = 1,25 \cdot 10^8 \text{ - для дзвінків;}$$

$$V_2^{R2} = 9,15 \cdot 10^8 \text{ - для sms-повідомлень;}$$

$$V_3^{R2} = 1,75 \cdot 10^9 \text{ - для mms-повідомлень;}$$

$$V_4^{R2} = 7,69 \cdot 10^8 \text{ - для Інтернет-контенту.}$$

Вирішимо аналогічну задачу для випадку, коли заявка проходить через інші функціональні блоки:

$$FB1.2 \rightarrow FB2.2 \rightarrow FB3 \rightarrow FB4.2 \rightarrow FB5$$

Кількість оперативної пам'яті, необхідної для обслуговування 1 заявки кожного типу.

$$S1: V_{R1}^1 = 1000 + 0 + 500 + 1000 + 200 = 2700$$

$$S2: V_{R1}^2 = 1000 + 1000 + 1000 + 100 + 1000 = 4100$$

$$S3: V_{R1}^3 = 1000 + 2000 + 1000 + 100 + 2000 = 6100$$

$$S4: V_{R1}^4 = 1000 + 500 + 600 + 100 + 400 = 2600$$

$$V_{R1} = 10^6 \cdot (V_{R1}^1 + V_{R1}^2 + V_{R1}^3 + V_{R1}^4) = 10^6 \cdot (2700 + 4100 + 6100 + 2600) = 1,55 \cdot 10^{10}$$

3) Загальна кількість постійної пам'яті

Кількість постійної пам'яті, необхідної для обслуговування 1 заявки кожного типу.

$$S1: V_{R2}^1 = 0 + 0 + 0 + 1000 + 0 = 1000$$

$$S2: V_{R2}^2 = 0 + 0 + 0 + 1000 + 0 = 1000$$

$$S3: V_{R2}^3 = 0 + 0 + 0 + 1000 + 0 = 1000$$

$$S4: V_{R2}^4 = 0 + 0 + 0 + 1000 + 0 = 1000$$

$$V_{R2} = 10^6 \cdot (V_{R2}^1 + V_{R2}^2 + V_{R2}^3 + V_{R2}^4) = 10^6 \cdot (1000 + 1000 + 1000 + 1000) = 4 \cdot 10^9$$

Таблиця 6.12

Кількість заявок кожного типу

k1	k2	k3	k4
1321102	880735	880731	917432

При цьому значення цільової функції, тобто максимального прибутку від обслуговування усіх заявок буде наступним:

$$\sum_{i=1}^4 k_i S_i = 16403662 (\text{y. o.})$$

Перевіримо виконання усіх необхідних умов.

- 1) $\sum_{i=1}^4 k_i (\sum_{j=1}^8 v_{ij}^{R1}) = 1321102 \cdot (1000 + 0 + 500 + 1000 + 200) + 880735 \cdot (1000 + 1000 + 1000 + 100 + 1000) + 880731 \cdot (1000 + 2000 + 1000 + 100 + 2000) + 917432 \cdot (1000 + 500 + 600 + 100 + 400) = 1,49 \cdot 10^{10} \leq V_{R1};$
- 2) $\sum_{i=1}^4 k_i (\sum_{j=1}^8 v_{ij}^{R2}) = 1321102 \cdot (0 + 0 + 0 + 1000 + 0) + 880735 \cdot (0 + 0 + 0 + 1000 + 0) + 880731 \cdot (0 + 0 + 0 + 1000 + 0) + 917432 \cdot (0 + 0 + 0 + 1000 + 0) = 4 \cdot 10^9 \leq V_{R2};$
- 3) $\frac{k_1}{k_2} = \frac{1321102}{880735} = 1,499999 \quad 0,5 \leq 1,499999 \leq 1,5;$
- 4) $\frac{k_1}{k_3} = \frac{1321102}{880731} = 1,500006 \quad 1 \leq 1,500006 \leq 2;$
- 5) $\frac{k_1}{k_4} = \frac{1321102}{917432} = 1,44 \quad 0,44 \leq 1,44 \leq 1,44;$
- 6) $\frac{k_2}{k_3} = \frac{880735}{880731} = 1,0000045 \quad 1 \leq 1,0000045 \leq 2;$
- 7) $\frac{k_2}{k_4} = \frac{880735}{917432} = 0,9600003 \quad 0,44 \leq 0,9600003 \leq 1,44;$

$$8) \frac{k_3}{k_4} = \frac{880731}{917432} = 0,9599959 \quad 0,125 \leq 0,9599959 \leq 1,125;$$

Тобто усі необхідні умови виконуються при даних значеннях k_i .

Визначимо необхідний об'єм ресурсів для знайденої кількості заявок кожного типу за формулами:

$$V_i^{R1} = k_i \sum_j v_{ij}^{R1} - \text{кількість оперативної пам'яті};$$

$$V_i^{R2} = k_i \sum_j v_{ij}^{R2} - \text{кількість постійної пам'яті}.$$

Отримаємо такі результати:

Таблиця 6.13

Розподіл ресурсів між заявками

R1	S1	S2	S3	S4
	3,57E + 09	3,6E + 09	5,4E + 09	2,38E + 09
R2	S1	S2	S3	S4
	1,32E + 09	8,8E + 08	8,8E + 08	9,17E + 08

Тобто кількість оперативної пам'яті для кожного типу сервісу дорівнює:

$$V_1^{R1} = 3,57 \cdot 10^9 - \text{для дзвінків};$$

$$V_2^{R1} = 3,6 \cdot 10^9 - \text{для sms-повідомлень};$$

$$V_3^{R1} = 5,4 \cdot 10^9 - \text{для mms-повідомлень};$$

$$V_4^{R1} = 2,38 \cdot 10^9 - \text{для Інтернет-контенту}.$$

Кількість постійної пам'яті для кожного типу сервісу дорівнює:

$$V_1^{R2} = 1,32 \cdot 10^9 - \text{для дзвінків};$$

$$V_2^{R2} = 8,8 \cdot 10^8 - \text{для sms-повідомлень};$$

$$V_3^{R2} = 8,8 \cdot 10^8 - \text{для mms-повідомлень};$$

$$V_4^{R2} = 9,17 \cdot 10^8 - \text{для Інтернет-контенту}.$$

Для того, щоб збільшити кількість заявок, які можуть обслуговуватись одночасно, потрібно збільшити кількість відповідних ресурсів, необхідних для їх обробки функціональними блоками. Для цього можна використовувати

додаткові віртуальні сервери у хмарах. Використання хмарних технологій дозволить збільшити об'єм необхідних ресурсів на сервері до будь-якого розміру.

Розглянемо на діаграмах, як буде змінюватись кількість заявок та відповідних ресурсів, необхідних для їх обслуговування при поступовому збільшенні об'єму ресурсів на сервері.

Будемо проводити дослідження залежності кількості заявок чотирьох різних типів та об'єму ресурсів для їх обслуговування, збільшуючи максимальний об'єм ресурсів оперативної та постійної пам'яті, який може бути доступним на сервері. Щоб визначити приріст ресурсів, який буде задаватись у процесі дослідження, визначимо необхідний об'єм оперативної та постійної пам'яті для потрібний функціональний блокам для обслуговування викликів, якщо на вхід сервера надходить два, три, чотири або п'ять мільйонів заявок кожного типу.

1) Для випадку $FB1.1 \rightarrow FB2.1 \rightarrow FB3 \rightarrow FB4.4 \rightarrow FB5$

Якщо $k_i=2000000$

$$V_{R1} = 2 \cdot 10^6 \cdot (V_{R1}^1 + V_{R1}^2 + V_{R1}^3 + V_{R1}^4) = 2,94 \cdot 10^{10}$$

$$V_{R2} = 2 \cdot 10^6 \cdot (V_{R2}^1 + V_{R2}^2 + V_{R2}^3 + V_{R2}^4) = 7,8 \cdot 10^9$$

Якщо $k_i=3000000$

$$V_{R1} = 3 \cdot 10^6 \cdot (V_{R1}^1 + V_{R1}^2 + V_{R1}^3 + V_{R1}^4) = 4,41 \cdot 10^{10}$$

$$V_{R2} = 3 \cdot 10^6 \cdot (V_{R2}^1 + V_{R2}^2 + V_{R2}^3 + V_{R2}^4) = 1,17 \cdot 10^{10}$$

Якщо $k_i=4000000$

$$V_{R1} = 4 \cdot 10^6 \cdot (V_{R1}^1 + V_{R1}^2 + V_{R1}^3 + V_{R1}^4) = 5,88 \cdot 10^{10}$$

$$V_{R2} = 4 \cdot 10^6 \cdot (V_{R2}^1 + V_{R2}^2 + V_{R2}^3 + V_{R2}^4) = 1,56 \cdot 10^{10}$$

Якщо $k_i=4000000$

$$V_{R1} = 5 \cdot 10^6 \cdot (V_{R1}^1 + V_{R1}^2 + V_{R1}^3 + V_{R1}^4) = 7,35 \cdot 10^{10}$$

$$V_{R2} = 5 \cdot 10^6 \cdot (V_{R2}^1 + V_{R2}^2 + V_{R2}^3 + V_{R2}^4) = 1,95 \cdot 10^{10}$$

Розрахуємо оптимальну кількість заявок кожного типу та об'єм ресурсів, необхідний для їх обслуговування для даних об'ємів оперативної та постійної пам'яті на сервері і зобразимо залежності у вигляді діаграм.

1) Залежність оптимальної кількості заявок чотирьох типів від збільшення максимальної кількості ресурсів сервера.

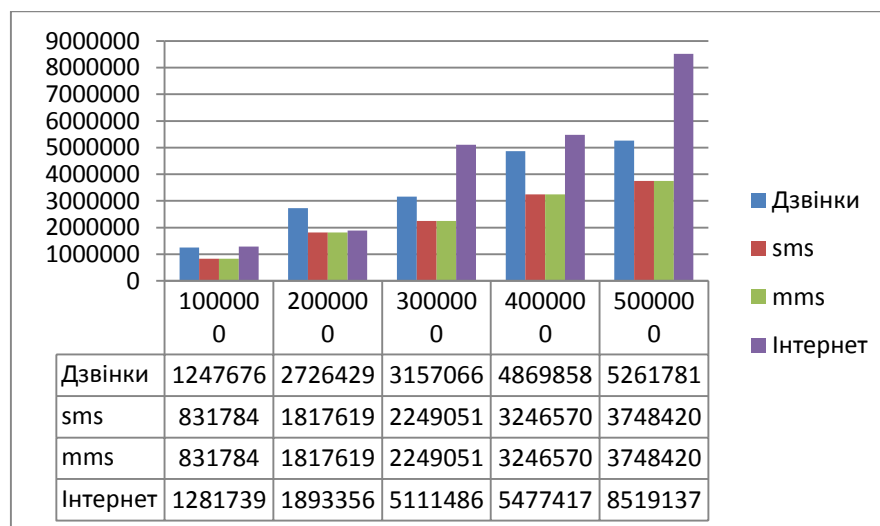


Рис. 6.11 Залежність кількості заявок від продуктивності сервера

2) Залежність об'єму оперативної пам'яті, який необхідно виділити для обслуговування кожного виду заявок від збільшення загальної кількості ресурсів на сервері.

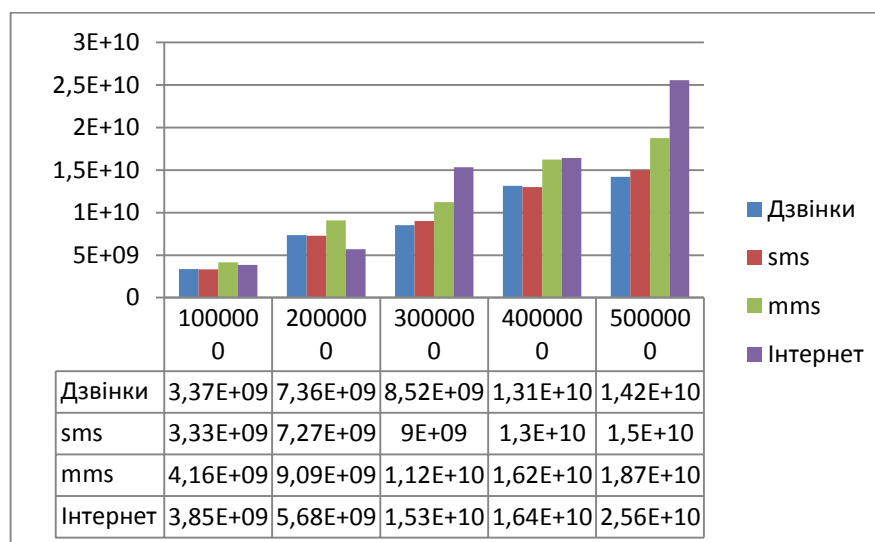


Рис. 6.12 Залежність кількості оперативної пам'яті для обслуговування кожного типу заявок від продуктивності сервера

3) Залежність об'єму постійної пам'яті, який необхідно виділити для обслуговування кожного виду заявок від збільшення загальної кількості ресурсів на сервері.

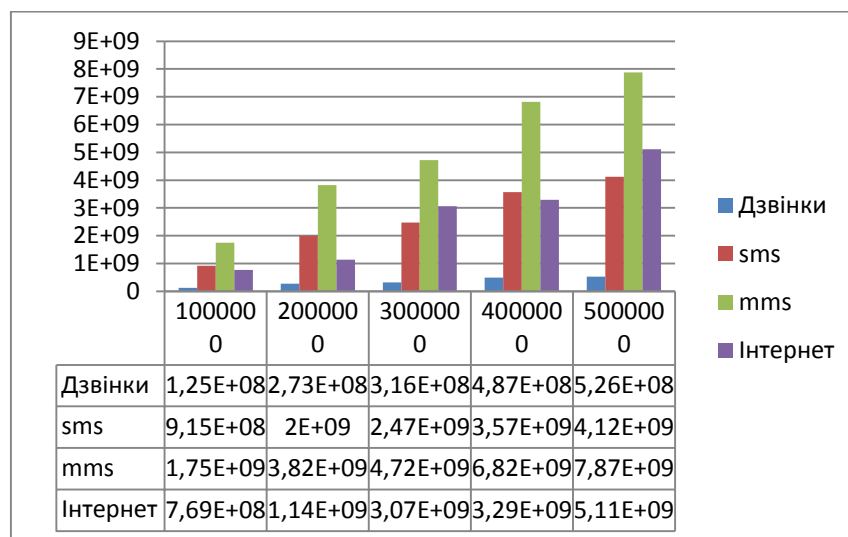


Рис. 6.13 Залежність кількості постійної пам'яті для обслуговування кожного типу заявок від продуктивності сервера

6.5. Імітаційна модель системи онлайн тарифікації із змінним ресурсом обслуговування

Імітаційна модель програмно-керованої мережі обчислювальних вузлів метою якої є створення інформаційно-обчислювальної мережі, яка динамічно організується на вимогу та виступає технічним засобом для вирішення задач системи тарифікації.

Розробити програмне забезпечення на основі запропонованих методів, що дозволяє розгорнути інформаційну мережу для забезпечення функціонування систем online тарифікації

Будується обчислювальна мережа, де

Вузол мережі – обчислювальна машина (віртуальна машина, яка створюється тільки для виконання однієї або групи задач системи online тарифікації (OCS).

Потоки мережі – інформаційні потоки, що генеруються в процесі обслуговування заявок на тарифікацію; керування потоками здійснюється централізовано за таблицями.

Таблиці керування потоками залежить від структури мережі, яка змінюється відповідно до розкладу.

Розклад включення серверів – залежить від дня року, окремо будується розклад для робочих та святкових днів.

Для не святкових днів розклад переключення записано в матриці $M1$ розмірністю $(N \times S)$, де N – кількість серверів, які планується підключати до системи в процесі обслуговування абонентів; $S=24*7*k$ – кількість переключень (зміна кількості серверів – учасників комутаційної мережі) за дубу та протягом тижня; k – кількість можливих переключень протягом години у не святковий день.

Для святкових днів розклад переключення серверів записано в матрицю $M2$ розмірністю $(N \times S1)$, де N – кількість серверів, які планується підключати до системи в процесі обслуговування абонентів; $S=24*k1$ – кількість переключень (зміна кількості серверів – учасників комутаційної мережі) за дубу; $k1$ – кількість можливих переключень протягом години у святковий день.

Ресурси вузла мережі – до них відноситься кількість оперативної пам'яті, кількість постійної пам'яті на дисках, потужність процесору, а також вхідний та вихідний каталі зв'язку. Ресурси вузла можуть бути змінені динамічно за розкладом розрахованим відповідно до запропонованого методу, моделювання методу наведено в п. 6.7.

Оцінювання якості обслуговування виконувалося на наборах статистичних даних, отриманих від компанії оператора зв'язку. У Додатку 1 наведено код програми, яка імітує обслуговування запитів, що надходять на вхід системи обслуговування, яка складається з набору вузлів (функціональних блоків). Ресурси кожного вузла визначаються конфігураціями та розкладом використання конфігурацій. На вхід подавалася

послідовність запитів для обслуговування, послідовність містила T_{mod} наборів запитів, кількість яких визначалася відповідно до отриманих від оператора зв'язку статистичних даних. На рисунку 19 наведено схему імітаційної моделі, на вхід подано.

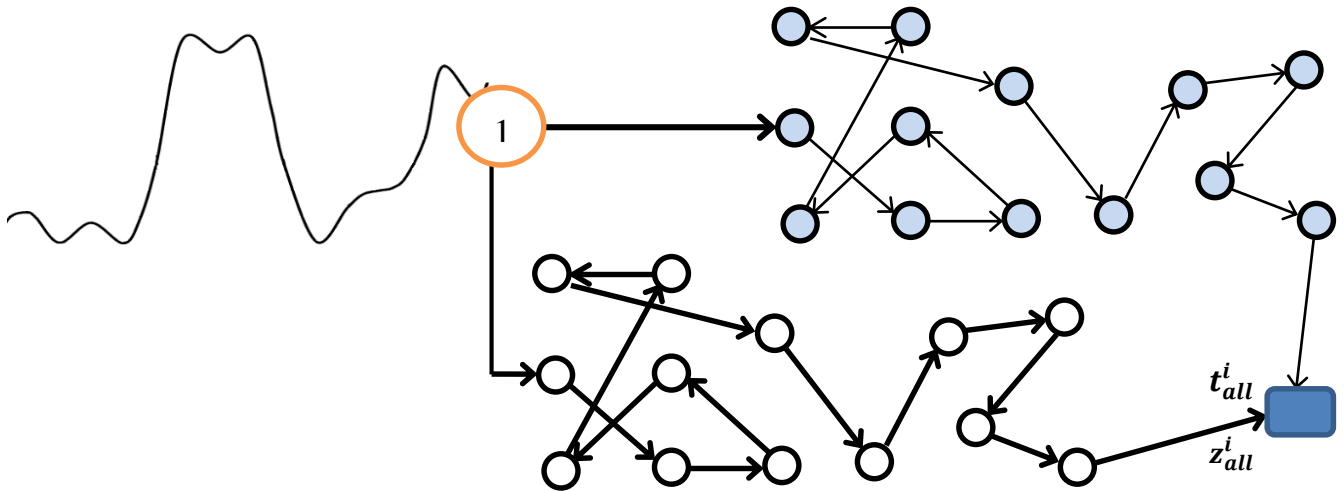


Рис. 6.14 – Імітаційна модель ГТС

Вхідні дані імітаційної моделі:

$dt = 0,1$ сек – інтервал дискретизації часу;

$T_{mod} = 864000$ інтервалів – загальний час моделювання (дискретний);

N_{Σ}^i – кількість запитів, які надійшли в систему за час i -го моделювання,

$i = \overline{1, 100}$;

$M = \begin{cases} 28, & \text{при викор. додаткових ресурсів;} \\ 14, & \text{інакше} \end{cases}$;

$V_j = \{V_{R1j}, V_{R2j}\}$ - доступні об'єми ресурсів вузла j .

Для кожного моделювання була зафіксована статистика показників якості обслуговування для кожного j -го вузла, а також в цілому по системі:

t_j^i – середній час затримки у вузлі j ($j = \overline{1, M}$), під час i -го моделювання
 $i = \overline{1, 100}$;

z_j^i – кількість втрачених запитів у вузлі j ($j = \overline{1, M}$), під час i -го моделювання
 $i = \overline{1, 100}$;

t_{all}^i - середній час затримки у системі під час i -го моделювання
 $i = \overline{1, 100}$;

z_{all}^i кількість втрачених запитів у системі під час i -го моделювання $i = \overline{1, 100}$.

Також системою моніторингу зберігалися дані про кількість ресурсів у кожний малий інтервал часу:

$R_{1j}^i = \{R_{1j}^1, \dots, R_{1j}^k, \dots, R_{1j}^{T_{\text{мод}}}\}$ – множина даних моніторинг ресурсу R1 у вузлі j ($j = \overline{1, M}$), під час i -го моделювання $i = \overline{1, 100}$;

$R_{2j}^i = \{R_{2j}^1, \dots, R_{2j}^k, \dots, R_{2j}^{T_{\text{мод}}}\}$ – моніторинг ресурсу R2 у вузлі j ($j = \overline{1, M}$), під час i -го моделювання $i = \overline{1, 100}$.

Для оцінки показників якості обслуговування гібридних сервісів виконувався розрахунок ймовірності того, що будуть порушуватися вимоги стандартів та специфікацій щодо для часу обслуговування та ймовірності вчасного обслуговування сервісу, відповідні формули для розрахунку ймовірностей було наведено у Таблиці 6.14. Результати моделювання зведені у Таблиці 6.15.

В роботі оператора зв'язку важливим показником функціонування системи в цілому є коефіцієнт використання ресурсів. Практика телекомунікаційної компанії показала, що коефіцієнт використання ресурсів повинен коливатися у межах від 30% до 80%. Оскільки, якщо коефіцієнт використання ресурсів більше 80% починають виникати непередбачувані збої, якщо коефіцієнт використання ресурсів менше 30%, тоді фіксується простій обладнання, та надлишкові витрати на його обслуговування. Тому в процесі моделювання оцінювалася ймовірність того, що ресурси системи використовуються менше заданого порогового значення a , а також ймовірність того що ресурси системи використовуються більше заданого порогового значення b

$$R_{2j}^i \supset R_{2j_a}^i = \{R_{2j}^k | R_{2j}^k < a * V_{R_{2j}}; k = \overline{1, T_{\text{мод}}}\} \quad |R_{2j_a}^i| = A^i \quad (20)$$

$$R_{2j}^i \supset R_{2j_b}^i = \{R_{2j}^k | R_{2j}^k > b * V_{R_{2j}}; k = \overline{1, T_{\text{мод}}}\} \quad |R_{2j_b}^i| = B^i \quad (21)$$

Таблиця 6.14 – Показники якості та відповідні оцінки якості

Показник якості	Порогове значення	Оцінка	Значення оцінки
t_3 - час затримки	$P_{tj}=0.8сек$	p_{1j} $\forall j = \overline{1, M}$	$p_{1j} = 1 - (\sum_{i=1}^{100} k_{ij})/100$ $k_{ij} = \begin{cases} 1 & t_j^i > P_{tj} \\ 0 & \text{інакше} \end{cases}$
	$P_{tall}=8 сек$	p_{1all}	$p_{1all} = 1 - (\sum_{i=1}^{100} K_{iall})/100$ $K_{iall} = \begin{cases} 1 & t_{all}^i > P_{tall} \\ 0 & \text{інакше} \end{cases}$
P – ймовірність успішного обслуговування	$P_{zi}=0,98$	p_{2j} $\forall j = \overline{1, M}$	$p_{2j} = 1 - (\sum_{i=1}^{100} K_{iz})/100$ $K_{iz} = \begin{cases} 1 & \frac{N_{\Sigma}^i - z_j^i}{N_{\Sigma}^i} > P_{zi} \\ 0 & \text{інакше} \end{cases}$
	$P_{zall}=0,98$	p_{2all}	$p_{2all} = 1 - (\sum_{i=1}^{100} K_{i2all})/100$ $K_{i2all} = \begin{cases} 1 & \frac{N_{\Sigma}^i - z_{all}^i}{N_{\Sigma}^i} > P_{zall} \\ 0 & \text{інакше} \end{cases}$
α – коефіцієнт використання ресурсів ГТС	$a=0,3$	$P_{3R_{2j}}$ $\forall j = \overline{1, M}$	$P_{3R_{2j}} = (\sum_{i=1}^{100} \frac{A_i}{T_{\text{мод}}})/100$
	$b=0,8$	$P_{4R_{2j}}$ $\forall j = \overline{1, M}$	$P_{4R_{2j}} = (\sum_{i=1}^{100} \frac{B_i}{T_{\text{мод}}})/100$

Таблиця 6.15 – Результати моделювання

	Стандартне обслуговування	Керування за методологією	Стандартне обслуговування	Керування за методологією
	Середня затримка обслуговуванні сервісу у вузлі ($\bar{t}_j = \sum_{i=1}^{100} t_j^i / 100$)		Оцінка вчасного обслуговування у вузлі ($p_1 = (\sum_{j=1}^M p_{1j})/M$)	

Max _j	7	9	0,8	0,805
Min _j	1	1		
Середнє	3,8	4,4		
	Середня затримка обслуговуванні сервісу по системі ($\overline{t_{all}} = \sum_i t_{all}^i / 100$)		Оцінка вчасного обслуговування ($p_2=p_{1all}$)	
max	80	90	0,82	0,84
Min	20	23		
Середнє	55	62		
	Ймовірність успішного обслуговування сервісу у вузлі ($\overline{z_j} = \sum_i \frac{N_{\Sigma}^i - z_j^i}{N_{\Sigma}^i} / 100$)		Оцінка ймовірності успішного обслуговування у вузлі ($p_3 = (\sum_{j=1}^M p_{2j}) / M$)	
Max _j	1	1	0,95	0.99
Min _j	0,96	0.99		
Середнє	0,98	0.999		
	Ймовірність успішного обслуговування сервісу по системі ($\overline{z_{all}} = \sum_i z_{all}^i / 100$)		Оцінка ймовірності успішного обслуговування по системі ($p_4 = p_{2all}$)	
Max _j	1	1	0,94	0.99
Min _j	0,95	0.99		
середнє	0,975	0.999		
	Коефіцієнт використання ресурсів ГТС α		Ймовірність використання ресурсу R1 і R2 з коефіцієнтом вик. обч. ресурсів менше заданого порогового значення ($\overline{P_{3R_{1j}}}$ і $\overline{P_{3R_{2j}}}$)	
(Max _j $\overline{R_{1j}}$)/ V_{R2j}	0.95	0.9	0.4 і 0,35	0,15 і 0,25
(Min _j $\overline{R_{1j}}$)/ V_{R2j}	0.15	0.25.	Ймовірність використання ресурсу R1 і R2 з коефіцієнтом вик. обч. ресурсів більше заданого порогового значення ($\overline{P_{4R_{1j}}}$, $\overline{P_{4R_{2j}}}$)	

$(\text{Max}_j \overline{R_{2j}}) / V_{R2j}$	1	0.85	0,2 і 0,15	0,05 і 0,1
$(\text{Min}_j \overline{R_{2j}}) / V_{R2j}$	0.1	0.2	Оцінка коефіцієнту вик. обч. ресурсів, які використовуються у межах норми $p_5 = 1 - \frac{\sum_{g=1}^2 w_g p_{3Rgj}}{2} - \frac{\sum_{g=1}^2 w_g p_{4Rgj}}{2}$	
середнє	0.4	0.2	0.43	0.75

Із таблиці 2 видно, що такі показники якості обслуговування як середня затримка обслуговуванні сервісу у вузлі і в цілому по системі, ймовірність успішного обслуговування сервісу у вузлі та в цілому по системі були утримані у межах допустимих значень та набули незначного покращення. Однак коефіцієнт використання ресурсів при обслуговуванні відповідно до запропонованих моделей та методів утримується у заданих межах із імовірністю більшою в середньому на 32%.

6.6. Імітаційна модель для методу вибору потужності обслуговуючого пристрою

Метод вибору потужності обслуговуючого пристрою для забезпечення процесу обслуговування на заданому рівні якості для багатоканального обслуговування [165].

Метод обслуговування заявок з використанням віртуалізованого обслуговуючого пристрою. Особливість застосування віртуалізованого обслуговуючого пристрою полягає у можливості збільшувати ресурси для виконання обчислювальних операцій. Така можливість з'являється через організацію віртуального обчислювального простору, який не прив'язаний до фізичного обладнання, але залежить від його кластерної організації. На відміну від звичайного підходу щодо балансування навантаження, коли для забезпечення зростаючого навантаження додається додаткові обслуговуючі пристрої між якими виконується розподілення запитів за деяким законом,

запропонований метод дозволяє при обслуговуванні телекомунікаційних сервісів, які характеризуються високою інтенсивністю вхідного потоку використовувати набір віртуальних сутностей, робота яких утворює єдиний простір обслуговування, та розрахувати параметри системи як єдиного обслуговуючого пристрою із змінною інтенсивністю обслуговування.

Для розрахунку рекомендованого значення потужності вузла обслуговування, необхідно обрати найбільш відповідні метрики хмарних модулів, підставити значення метрик у модель і провести розрахунок значення потужності вузла обслуговування.

- λ – середнє значення проміжків часу між моментами надходження замовлень,

- n – кількість заявок, які можуть одночасно обслуговуватися у віртуальному вузлі.

- v_k – кількість узагальненого обчислювального ресурсу, який використовується для обслуговування однієї заявки при одночасному обслуговуванні k заявок $k = \overline{1, n}$.

- a, b – верхня та нижня границя завантаженості системи обслуговування,

- R – задане експертами значення, який показує долю часу роботи системи без черги.

V – кількість узагальненого ресурсу

В Таблиці 6.16 приведено значення, що були взяті на системі обслуговування хмарних обчислень в момент найбільшої завантаженості.

Таблиці 6.16 Вхідні данні моделі

Параметр	Одночасне значение
n	500
λ	648
v_k^g	4Mb
V^g	1612Mb
R	90%
a	0.3

b	0.8
---	-----

Програмна реалізація моделі.

Програмна реалізація виконана за допомогою програмної мови Python. Python - це інтерпретована мова програмування високого рівня для програмування загального призначення. Python має філософію дизайну, яка підкреслює читаність коду, особливо використовуючи значні прогалини. Він надає конструкції, які забезпечують чітке програмування як на малих, так і на великих масштабах.

Файл `calc_mu.py` призначений для розрахунку імовірностей та визначення цільової функції r .

Лістинг 6.7 відображає програмну реалізацію розрахунку потужності вузла обслуговування. В даній програмі рналізовано обрану математичну модель, значення інтенсивності обробки вхідного потоку заявок розрахованого відповідно до запропонованого методу.

Результат пошуку потужності вузла хмарних обчислень для вказаного прикладу відображено на Рисунку 6.15.

```

r(l=648, mu=989.00, cel=2.6208291203235583) = 0.9255416024328514
r(l=648, mu=989.50, cel=2.6195048004042443) = 0.9249041996794265
r(l=648, mu=990.00, cel=2.6181818181818177) = 0.9242676514439666
r(l=648, mu=990.50, cel=2.6168601716304893) = 0.9236319560439881
r(l=648, mu=991.00, cel=2.6155398587285577) = 0.9229971118013489
r(l=648, mu=991.50, cel=2.614220877458396) = 0.9223631170422353
r(l=648, mu=992.00, cel=2.612903225806452) = 0.9217299700971481
r(l=648, mu=992.50, cel=2.611586901763224) = 0.9210976693008897
r(l=648, mu=993.00, cel=2.6102719033232633) = 0.9204662129925482
r(l=648, mu=993.50, cel=2.6089582284851542) = 0.9198355995154865
r(l=648, mu=994.00, cel=2.6076458752515093) = 0.9192058272173276
r(l=648, mu=994.50, cel=2.6063348416289593) = 0.9185768944499395
r(l=648, mu=995.00, cel=2.6050251256281407) = 0.9179487995694247
r(l=648, mu=995.50, cel=2.603716725263687) = 0.9173215409361055
r(l=648, mu=996.00, cel=2.6024096385542173) = 0.9166951169145093
r(l=648, mu=996.50, cel=2.601103863522328) = 0.9160695258733574
r(l=648, mu=997.00, cel=2.599799398194584) = 0.9154447661855512
r(l=648, mu=997.50, cel=2.5984962406015035) = 0.9148208362281575
r(l=648, mu=998.00, cel=2.597194388777555) = 0.9141977343823974
r(l=648, mu=998.50, cel=2.5958938407611414) = 0.9135754590336321
r(l=648, mu=999.00, cel=2.5945945945945956) = 0.9129540085713501
r(l=648, mu=999.50, cel=2.5932966483241624) = 0.912333381389154
r(l=648, mu=1000.00, cel=2.5920000000000001) = 0.9117135758847482
r(l=648, mu=1000.50, cel=2.5907046476761613) = 0.9110945904599248
r(l=648, mu=1001.00, cel=2.589410589410588) = 0.9104764235205524
r(l=648, mu=1001.50, cel=2.5881178232651023) = 0.9098590734765611
!!!
Process finished with exit code 0

```

Рис. 6.15 Результат розрахунку файлу `calc_mu.py`

Лістинг 6.7 Реалізація розрахунку потужності вузла обслуговування

```

import numpy as np

N = 500 # кількість заявок, які можуть одночасно обслуговуватись у віртуальному вузлі
V = 1612 # кількість узагальненого ресурсу
VK = 4 # кількість ресурсу, необхідного для обслуговування однієї заявки
A = .3 # нижня границя завантаженості системи обслуговування
B = .8 # верхня границя завантаженості системи обслуговування

MU_STEP = 0.5 # Точність визначення інтенсивності обслуговування

def calc_speed(k, lmbd, mu):
    return lmbd / (k * mu)

def calc_pa(n, lmbd, mu):
    pa = 1
    for i in range(1, n + 1):
        pa *= calc_speed(i, lmbd, mu)
    return pa

def calc_p0(n, lmbd, mu):
    epa = 0
    for k in range(1, n+1):
        epa += calc_pa(k, lmbd, mu)

    return (1 + epa + (calc_speed(n, lmbd, mu) / (1 - calc_speed(n, lmbd, mu))) * calc_pa(n, lmbd, mu)) ** -1

def calc_pk(k, p0, lmbd, mu):
    return p0 * calc_pa(k, lmbd, mu)

def calc_pnk(n, p0, kk, lmbd, mu):
    return p0 * calc_speed(n, lmbd, mu) ** kk * calc_pa(n, lmbd, mu)

def calc_r(n, mu, lmbd):
    r = 0
    p0 = calc_p0(n, lmbd, mu)
    for k in range(1, n+1):
        r += calc_pk(k, p0, lmbd, mu) / p0

    return r

def calc_v(n, mu, lmbd):
    v = 0
    p0 = calc_p0(n, lmbd, mu)
    for k in range(1, n+1):
        v += VK * k * calc_pk(k, p0, lmbd, mu) / p0

    return v

def cel_f(n, mu, lmbd):
    cel = 0
    p0 = calc_p0(n, lmbd, mu)
    for k in range(1, n+1):
        cel += k*VK*calc_pk(k, p0, lmbd, mu)
    return cel

def main():
    with open('out.txt', 'w') as f:
        l=648 # середнє значення проміжків часу між моментами надходження замовлень
        for mu in np.arange(1 / B, 1 / A + MU_STEP, MU_STEP):
            rr = calc_r(N, mu, l)
            vv = calc_v(N, mu, l)
            cel=cel_f(N, mu, l)
            res = f"r(l={l}, mu={mu:.2f}, cel={cel}) = {rr}"
            print(res)
            if .89 < rr < .91 and vv <= V:
                f.write(res + '\n')
                print('\t!!!')
                break

if __name__ == '__main__':
    main()

```


При даних обмеженнях і цільовій функції обчислене значення $\mu=1001.5$ При цьому сама цільова функція має значення 2.58811.

При збільшенні λ значення μ також збільшується, що підтверджує працездатність обраного методу.

Розрахунок оптимального значення інтенсивності вхідного потоку для існуючої системи

Для розрахунку оптимального значення λ проведено аналіз доступних у хмарній системі метрик для найнавантаженого дня у 2017 році (24.11.2017 «Чорна п'ятниця»). Максимальне значення інтенсивності вхідного потоку було отримано у години найбільшого навантаження, на момент, коли метрики сигналізували про потребу у додатковому ресурсі.

Під час аналізу, для кожного запису була розрахована інтенсивність вхідного потоку (λ) і одночасне значення інтенсивності обслуговування заявки (μ). Базуючись на отриманих даних, можна зробити висновок, що максимальному допустимому значенню інтенсивності вхідного потоку для єдиного процесу, при якому не виникає деградації сервісу (відсутність відмов) відповідає $\lambda = 648$ при $\mu = 1001.5$

Дані розрахунку оптимального значення λ і μ занесені в Таблицю 6.17.

Таблиця 6.17 Дані розрахунку оптимального значення λ і μ

Дата	Час	μ	λ
24.11.2017	00:00	930.5	601.93
	01:00	621.5	401.95
	02:00	587	379.95
	03:00	575	371.96
	04:00	566	365.96
	05:00	603	389.95
	06:00	692.5	447.95
	07:00	850	549.93
	08:00	1375.5	889.89
	09:00	1227.5	793.90
	10:00	1283	829.90
	11:00	1323	855.89
	12:00	1351	873.89
	13:00	1360	879.89

	14:00	1338.5	865.89
	15:00	1331.5	861.89
	16:00	1286	831.90
	17:00	1329	859.89
	18:00	1338,5	865.89
	19:00	1314	849.89
	20:00	1273.5	823.90
	21:00	1131.5	731.91
	22:00	1001.5	647.92
	23:00	859.5	555.93

Графічне відображення розрахованої інтенсивності обслуговування заявки (μ) зображено на графіку 4.1, а графічне відображення одночасного значення інтенсивності обслуговування заявки (μ) і інтенсивності вхідного потоку (λ), для єдиного процесу, зображено на графіку 4.2.



Графік. 4.1 Графік значення інтенсивності μ

Розрахунок значення інтенсивності обслуговування вхідного потоку заявок для усунення втрат, в системі, де наявні втрати.

Для повноти аналізу проведемо розрахунок інтенсивності обробки потоку вхідних заявок в системі, в якій наявні відмови, для їх усунення.

В Таблиці 4.18 чітко видно що інтенсивність обробки вхідного потоку заявок, що доступна системі (μ) значно менша за ту, що пропонує

використаний метод ($\mu_{\text{роз}}$), однак, це обумовлено тим що μ розрахована при умові допустимості черги та відмов, а $\mu_{\text{роз}}$ розрахована так щоб гарантовано обслужити всі заявки та мати запас 10% (при $R=90\%$). Для систем хмарних обчислень така така принята доля часу роботи без черги є актуальною, адже при затримках у отриманні провайдером даних з сервісу, спричиненими мережевими затримками, необхідно мати час на прорахунок необхідного значення $\mu_{\text{роз}}$ та надсилання керуючого сигналу до моменту початку відмов.

Однак, слід зазначити що $\mu_{\text{роз}}$ – найменше можливе значення для обраного методу і при її збільшенні фактична доля часу роботи системи без черги зменшується.

Таблиця 4.18 Данні порівняльного розрахунку μ і $\mu_{\text{роз}}$

Дата	Час	Передбачувані відмови при μ	μ	λ	$\mu_{\text{роз}}$
11.04.2018	00:00	0	323	645.92	998.5
	01:00	0	217	433.95	671
	02:00	0	205	409.95	634
	03:00	0	200	399.95	618
	04:00	0	198	395.95	612
	05:00	0	211	421.95	652
	06:00	0	241	481.94	744
	07:00	0	296	591.93	915
	08:00	100	478	955.88	1477.5
	09:00	100	427	853.89	1320
	10:00	300	446	891.89	1378.5
	11:00	300	460	919.89	1421
	12:00	400	471	941.88	1456
	13:00	450	474	947.88	1464
	14:00	400	466	931.88	1440.5
	15:00	300	456	911.89	1409
	16:00	300	448	895.89	1385
	17:00	300	449	897.89	1388
	18:00	300	446	891.89	1378.5
	19:00	200	438	875.89	1354
	20:00	0	423	845.90	1307.5
	21:00	0	395	789.90	1221
	22:00	0	350	699.91	1082
	23:00	0	300	599.93	927.5

Графічне відображення розрахованої інтенсивності обслуговування заявок ($\mu_{роз}$), доступної інтенсивності обслуговування заявок (μ), відмов при цьому значенні доступної інтенсивності обслуговування заявок і одночасного значення інтенсивності вхідного потоку (λ), для наявної системи зображено на Рис.16.6.

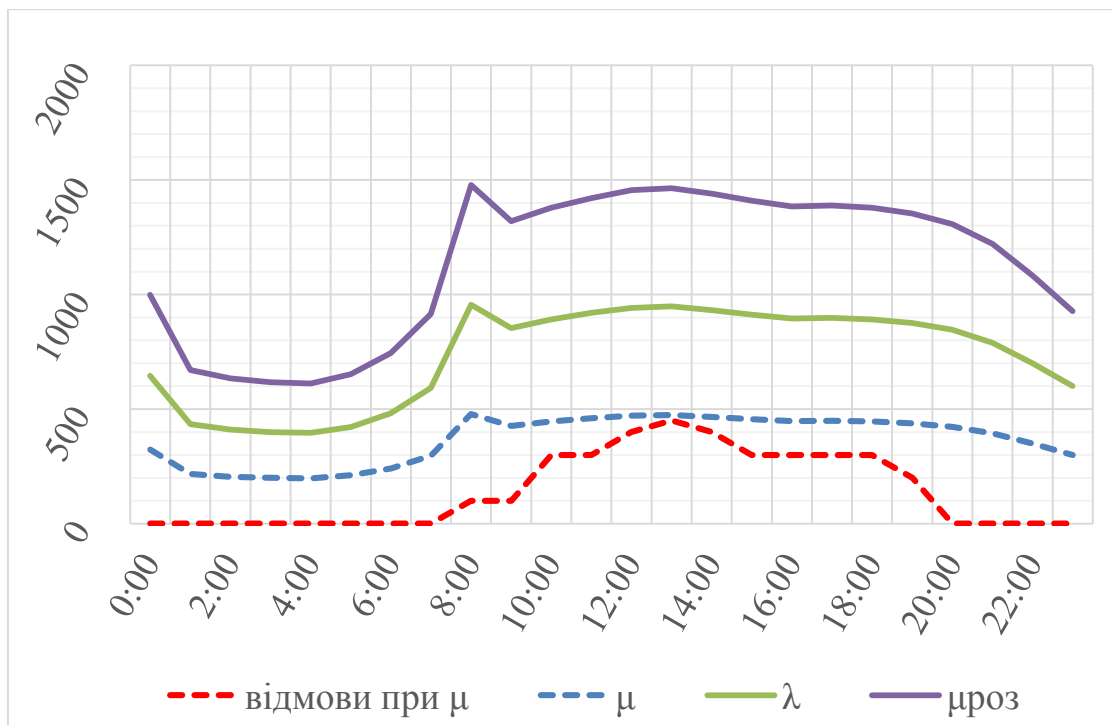


Рис 6.16 Графік $\mu_{роз}$ для запропонованої системи

Розрахунок прогнозу показує значне підвищення розрахованої інтенсивності обслуговування заявок ($\mu_{роз}$) при скачках інтенсивності вхідного потоку заявок (λ), що гарантує роботу системи без черг та відмов.

6.7. Імітаційна модель MATLAB для методу побудови розкладу залучення ресурсів на основі довгострокової статистики із забезпеченням показників якості та енергоефективності

Імітаційна модель методу побудови розкладу залучення ресурсів на основі довгострокової статистики складається з 6 файлів.

1. Файл `res.m` призначений для розрахунку імовірностей та визначення цільової функції `g`.

```
function [g]=res(x)
global l % порогова кількість запитів у черзі на обслуговування
global m % максимальна кількість запитів у черзі на обслуговування
global p % масив імовірностей
global lambda % значення інтенсивності вхідного потоку
% x(1)=mu - інтенсивність обслуговування заявки

alpha=lambda/4/x(1);

A=1-alpha^(m-l)-(m-l)*(1-alpha)*alpha^(m+2);
p(1)=(1-alpha)*(1-alpha^(m-l))/A; % значення p0, оскільки нумерація з 1
for i=2:1:(l+2)
    p(i)=alpha^(i-1)*p(1); % значення pk, 1<=k<=l+1
end
for i=(l+3):1:(m+1)
    p(i)=(1-alpha)*(alpha^(i-1)-alpha^(m+1))/A; % значення pk, l+2<=k<=m
end
p(m+2)=(1-alpha)^2*alpha^(m+1)/A;
g=x(1); % цільова функція
```

2. Файл `rosenbrock.m` призначений для надання всіх необхідних констант та ініціалізації робочих масивів. В ньому слід задати (записати в текстовому редакторі) наступні константи:

$\mu^k = (4, 8, 12, 16, 20)$ – інтенсивність обслуговування заявок за умови k -ї конфігурації обслуговуючого пристрою, що забезпечує показники якості обслуговування на заданому рівні, тобто кількість заявок за одиницю часу, які можуть бути обслужені, $k=1,..K$, K – кількість можливих конфігурацій системи обслуговування. –

λ - середнє значення проміжків часу між моментами надходження замовлень, незалежні випадкові величини, розподілені за показниковим законом, дані вводяться з файлу 1.xls,

$\varepsilon=0,05$ - мале число

$\Delta = 0,01\text{с}$ - Допустимий час затримку заявок у процесі обслуговування. –

$l = 4$ – порогове значення кількості заявок у черзі –

$m = 5$ – максимальна кількість заявок у черзі. –

$s_1=1, s_2=1, s_3=2, s_4=2, s_5=2$ для етапа 3

$P_{\text{lost}}=0,05$ – допустимий відсоток втрачених пакетів.

```
function h = rosenbrock(x)
```

```
global K    % Кількість конфігурацій
```

```
global eps  % похибка
```

```
global l    % порогова кількість запитів у черзі на обслуговування
```

```
global m    % максимальна кількість запитів у черзі на обслуговування
```

```
global Plost % допустима частка втрачених пакетів
```

```
global R    % частка заявок, які обслуговуються у системі не більше доп. часу
```

```
global rho  % оптимальне завантаження системи
```

```
global s    % кількість значень імовірностей для розрахуку обмежень
```

```
global p    % масив імовірностей
```

```
% задаємо значення параметрів
```

```
K=5;
```

```
eps=0.05;
```

```
l=4;          % порогова кількість запитів у черзі на обслуговування
```

```
m=5;          % максимальна кількість запитів у черзі на обслуговування
```

```
p=zeros(1,m+2);
```

```
Plost=0.05;   % допустима частка втрачених пакетів
```

```
R=0.8;        % відсоток заявок, які обл.у системі не більше доп. часу
```

```
rho=0.8;      % оптимальне завантаження системи
```

```
s=[1 1 2 2 2]; % кількість значень імовірностей для розрахуку обмежень
```

```
h=res(x);
```

3. Файл `unitdisk.m` призначений для формування обмежень на кількість ресурсів обслуговування.

```
function [c, ceq] = unitdisk(x)
global K    % Кількість конфігурацій
global mu   % інтенсивність обробки при K-й конфігурації
global Plost % допустима частка втрачених пакетів
global R    % відсоток заявок, які обслуговуються у системі не більше доп. часу
global rho   % оптимальне завантаження системи
global s     % кількість значень імовірностей для розрахунку обмежень
global m     % максимальна кількість запитів у черзі на обслуговування
global p     % масив імовірностей
global lambda % значення інтенсивності вхідного потоку
% x(1) - номер конфігурації

P=0;
a=floor(x(1));    % ціла частина від номера конфігурації
for i=1:1:s(a)
    P=P+p(i+1);    % сума імовірностей
end
mu=(4*x(1));      % інтенсивність обробки при K-й конфігурації
% Формування вектору неточних обмежень
c(1,1)=lambda/mu-rho; % обмеження на навантаження системи
c(1,3)=p(m+2)-Plost;  % обмеження на частку втрачених пакетів
c(1,3)=P-R;           % обмеження на частку заявок, обсл. у системі
c(1,4)=x(1)-K;        % обмеження на кількість конфігурацій згори
c(1,2)=-x(1)+1;       % обмеження на кількість конфігурацій знизу
ceq = [];             % точні обмеження відсутні
```

4. Файл `func.m` призначений для розрахунку функції $f(t)$.

```
function func()
```

```

global data    % масив вхідних даних
global L       % кількість інтервалів усереднення в функції
global DT      % інтервал усереднення
global f       % апроксимуюча функція
data=xlsread('1.xls'); % читаємо файл навчальних даних
n=length(data);
DT=100;        % інтервал усереднення
K=floor(n/DT);
for k=1:1:DT
    t(k)=k;    % формуємо масив часу
end;
j=1;
for k=1:1:K
    for i=1:1:DT
        x(i)=data(j); % формуємо масив даних на інтервалі усереднення
        j=j+1;
    end
    f(k,1)=(k-1)*DT+1; % початок інтервалу усереднення
    f(k,2)=k*DT;      % кінець інтервалу усереднення
    [a]=polyfit(t,x,1); % лінійна апроксимація МНК
    f(k,3)=a(1);      % нахил
    f(k,4)=a(2);      % зміщення
end
% fprintf('%5.0f %5.0f %3.5f %3.5f\n',f');
% об'єднання інтервалів з близькими показниками апроксимації
eps=0.05;
k=2;
ll=0;
L=K;
while k<L
    if (abs(f(k,3)-f(k-1,3))<=eps) && (abs(f(k,4)-f(k-1,4))<=eps)
        % якщо треба об'єднати інтервали
    end
end

```



```

f(k-1,2)=f(k,2);          % формуємо попередній рядок f
f(k-1,3)=(f(k-1,3)+f(k,3))/2;
f(k-1,4)=(f(k-1,4)+f(k,4))/2;
L=K-11;
for i=(k+1):1:L
    f(i-1,1)=f(i,1);      % підтягуємо f
    f(i-1,2)=f(i,2);
    f(i-1,3)=f(i,3);
    f(i-1,4)=f(i,4);
end
for i=(K-11):1:K
    f(i,1)=0;             % обнуляємо зайве f
    f(i,2)=0;
    f(i,3)=0;
    f(i,4)=0;
end
11=11+1;
k=k-1;
end
% якщо не треба об'єднати інтервали
k=k+1;
end
%fprintf('%5.0f %5.0f %3.5f %3.5f\n',f);

```

5. Файл `lambda1.m` призначений для розрахунку параметрів вхідних даних.

```

function lambda1(k)
global data    % масив вхідних даних
global L      % кількість інтервалів усереднення в функції
global lambda % розраховане значення для інтенсивності вхідного потоку
global sigma  % середньоквадратичне відхилення
global f      % апроксимуюча функція

```

```

global N      % кількість перемикачів
global nk     % номер інтервалу, на якому треба зробити перемикання
for i=1:1:N
    y(1,i)=i;
    y(2,i)=data(N*(k-1)+i);
end;
sigma=zeros(1,L-1);
for j=1:1:(L-1)
    a=f(j,2)-f(j,1)+1;
    if a<N N=a; N=N; end
    sig=0;
    for i=1:1:N
        sig=sig+(y(2,i)-f(j,3)*i-f(j,4))^2;
    end
    sigma(j)=sqrt(sig/(N-1));
end
nk=find(sigma==min(sigma));
sigma=min(sigma);
lambda=0;
for j=1:1:N
    lambda=lambda+y(2,j);
end
lambda=lambda/N;

```

6. Файл start.m призначений для запуску програми на виконання.

Використовується функція оптимізації fmincon.

```

global data    % масив вхідних даних
global sigma   % середньоквадратичне відхилення
global lambda  % розраховане значення для інтенсивності вхідного потоку
global nk      % номер інтервалу, на якому треба зробити перемикання
global f       % апроксимуюча функція

```

```

global N      % тривалість інтервалу вимірювання
func         % побудова навчальної функції
%options = optimset('Display','iter','Algorithm','active-set');
n=length(data);
N=100;
K1=floor(n/N);
% K1=5;
swich=zeros(K1,4);
for k=1:1:K1
lambda1(k);    % значення інтенсивності вхідного потоку
[x,fval] = fmincon(@rosenbrock,[1],...
                  [],[],[],[],[],@unitdisk,options);

k1=floor(x(1)); % конфігурація, на яку треба перейти
mu=4*k1;       % інтенсивність обслуговування в k-й конфігурації
mu0=4*(k1-1);  % інтенсивність обслуговування в k-й конфігурації
N=f(nk,2)-f(nk,1)+1;
%N=20;
for i=1:1:N
    if (mu-(f(nk,3)*i+f(nk,4))*4)<3*sigma break,end;
end
tk=((k-1)*N+i)*5;
swich(k,1)=lambda;
swich(k,2)=sigma;
swich(k,3)=tk;
swich(k,4)=mu;
end
%fprintf('%5.0f %5.0f %3.5f %3.5f\n',f');
fprintf('lambda sigma  tk    mu\n');
fprintf('%5.3f %5.3f %5.0f %5.0f\n',swich);
fprintf('\n');

```

Для запуску програми слід переписати всі вказані файли в робочу папку Matlab і запустити файл start.m.

Результат розрахунку для вказаного прикладу (наведено тільки кінцеві результати):

```
>> start
```

lambda	sigma	tk	mu
0.870	0.636	35	4
0.320	0.598	1000	4
0.400	0.651	1500	4
0.180	0.405	2000	4
0.070	0.260	7500	4
6.110	2.061	2505	4
0.020	0.141	10500	4
7.050	2.403	3505	8
0.180	0.386	9000	4
4.480	2.338	4505	4
1.770	1.262	5005	4
2.880	1.767	5505	4
5.160	2.243	6005	4
5.340	1.652	6505	4
6.330	1.984	7005	4
6.110	2.061	7505	4
5.860	2.317	8005	4
5.670	2.146	8505	4
4.480	2.338	9005	4
7.080	2.095	9505	8
7.700	2.107	10005	8
7.050	2.403	10505	8
5.420	1.621	11005	4
4.260	1.736	11505	4
4.270	1.393	12005	4
4.060	1.376	12505	4
1.970	1.187	13005	4
3.200	1.459	13505	4

Отримано масив моментів переключення та потрібних конфігурацій для кожного λ . Перевірка розрахованого розкладу у імітаційній моделі наведених у першій частині Додатку 1 за наборами статистичних даних наведених у другій частині Додатку 1 показала, що для кожної статистичної вибірки обчислювальних ресурсів відповідно до імітаційної моделі було достатньо для обслуговування запитів без додаткової затримки.

6.8. Імітаційна модель MATLAB для методу короткострокового прогнозування навантаження

Програма для розрахунку складається з 3 файлів.

1. Файл unitdisk.m призначений для формування обмежень на кількість ресурсів обслуговування.

λ_i - кількість заявок за 1мс, ($i \in 0..N$), $N = T_{inf} / 1мс$, $\lambda_i \in \Lambda$, Λ - множина значень статистики кількості заявок, що надходили протягом часу T_{inf} до початку здійснення прогнозу, $|\Lambda| = N$.

```
function [c, seq] = unitdisk(x)
global P      % імовірність помилки прогнозування
global p      % розрахована помилка прогнозування
global p2     % розрахована помилка прогнозування
c=p-1+P;      % обмеження на помилку прогнозування
seq = [p2];   % точні обмеження
```

2. Файл rosenbrock.m призначений для імовірностей та визначення цільової функції g.

```

function [g] = rosenbrock(x)
global data          % масив вхідних даних
global Tn            % інтервал часу, для якого потрібен прогноз
global Tnp           % період прогнозування
global M             % гранична кількість заявок
global P             % імовірність помилки прогнозування
global p             % розрахована помилка прогнозування
global p2            % розрахована помилка прогнозування

Tinf=floor(x(1));    % час збирання інформації
if Tinf>0
DeltaT=Tn(2)-Tn(1); % тривалість часу, для якого потрібен прогноз
Z=0;                % лічильник прапорців перемикавання
for j=1:(DeltaT-Tinf)
for k=1:Tinf
t(k)=k;             % формуємо масив часу
y(k)=data(Tn(1)+j-1+k); % формуємо масив даних на інтервалі усереднення
end;

[a]=polyfit(t,y,1); % лінійна апроксимація МНК
sig=0;
for k=1:Tinf
sig=sig+(y(k)-a(1)*k-a(2))^2;
end
sigma=sqrt(sig/(Tinf-1));
lambda=a(1)*Tnp+a(2)+3*sigma;
if lambda>M
z=1;
else
z=0;
end
Z=Z+z;              % нараджуємо лічильник прапорців перемикавання
end
p=Z/(DeltaT-Tinf); % розрахована помилка прогнозування
g=x(1);            % цільова функція
fprintf ('%5.2f %5.2f\n',Tinf,p-1+P);
else
p2=0;
g=0;
end
end

```

3. Файл start.m призначений для запуску програми на виконання та надання всіх необхідних констант Використовується функція оптимізації fmincon.

В ньому слід задати (записати в текстовому редакторі) наступні константи:

- $M=2$ – гранична кількість заявок, яка може бути обслужена при заданій конфігурації обслуговуючого пристрою.
- $P=0,1$ – імовірність помилки прогнозування.

- $T_{пр} = 10\text{мс}$ - період прогнозування, час, по завершенню якого запускається алгоритм прогнозування.

Слід відзначити, що алгоритм пошуку мінімального значення $T_{інф}$ дуже залежить від початкових значень для ітерації. Тому доцільно спочатку зробити грубу оцінку $T_{інф}$, використовуючи перебір можливих значень $T_{інф}$ (закоментований фрагмент), а потім задати максимальне значення з обраного діапазону та запустити програму `fmincon`.

```
global data % масив вхідних даних
global Tn % інтервал часу, для якого потрібен прогноз
global Tnp % період прогнозування
global M % гранична кількість заявок
global P % імовірність помилки прогнозування
global p2 % розрахована помилка прогнозування
% задаємо значення параметрів
M=2; % гранична кількість заявок
P=0.1; % імовірність помилки прогнозування
Tn=[500 2000]; % початкове значення вхідної послідовності
Tnp=10; % період прогнозування
Z=0; % лічильник прапорців перемикавання
p2=1;

options = optimset('Display','iter','Algorithm','interior-point');
data=xlsread('1.xls'); % читаємо файл навчальних даних
n=length(data);

[x,fval] = fmincon(@rosenbrock,[570],...
    [],[],[],[],[],[],[],@unitdisk,options);
%for i=100:50:1000
%rosenbrock(i);
%end
x(1)
```

Результат виконання програмної реалізації вирішення поставленої задачі визначив оптимальний $T_{інф}$ для досліджуваних даних.

Iter	F-count	f(x)	Feasibility	optimality	step
0	2	5.700000e+002	1.000e+000	1.000e+000	
1	4	5.692224e+002	1.000e+000	1.000e+000	7.776e-001
2	10	5.667463e+002	1.000e+000	1.000e+000	2.476e+000
3	13	5.666173e+002	1.000e+000	1.000e+000	1.289e-001
4	15	5.659660e+002	1.000e+000	1.000e+000	6.513e-001
5	21	5.636402e+002	1.000e+000	1.000e+000	2.326e+000
6	27	5.469985e+002	1.000e+000	1.000e+000	1.664e+001
7	30	5.468728e+002	1.000e+000	1.000e+000	1.258e-001
8	32	5.459751e+002	1.000e+000	1.000e+000	8.976e-001
9	34	5.388083e+002	1.000e+000	1.000e+000	7.167e+000

Local minimum possible. Constraints satisfied.

ans =

5.388083e+002

Таким чином, значення оптимального $T_{\text{інф}}$, при якому для даного прикладу виконуються всі умови, складає 538.

Результат роботи запропонованого методу з оптимальним $T_{\text{інф}}$

В експерименті було вибрано значення $T_{\text{інф}}=538$. Звернувши увагу на прогнозовані дані (рис 6.16) та реальні (рис 6.17), можна виділити наступний висновок: результат роботи запропонованого алгоритму відповідає реальним даним.

Після більш детальної оцінки результатів було виявлено, що:

1. 80% відсотків спрогнозованих даних виявились вірними.
2. В 95% відсотках запропонований метод вдало запропонував змінити конфігурацію обладнання.
3. Час виконання алгоритму склав 25 секунд.

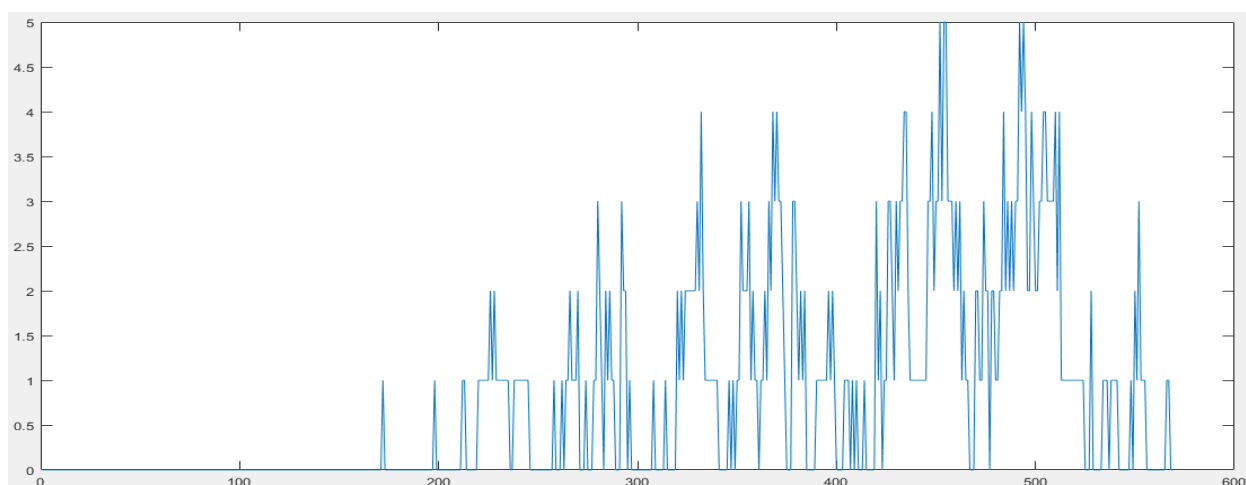


Рис 6.16 Результат роботи запропонованого методу

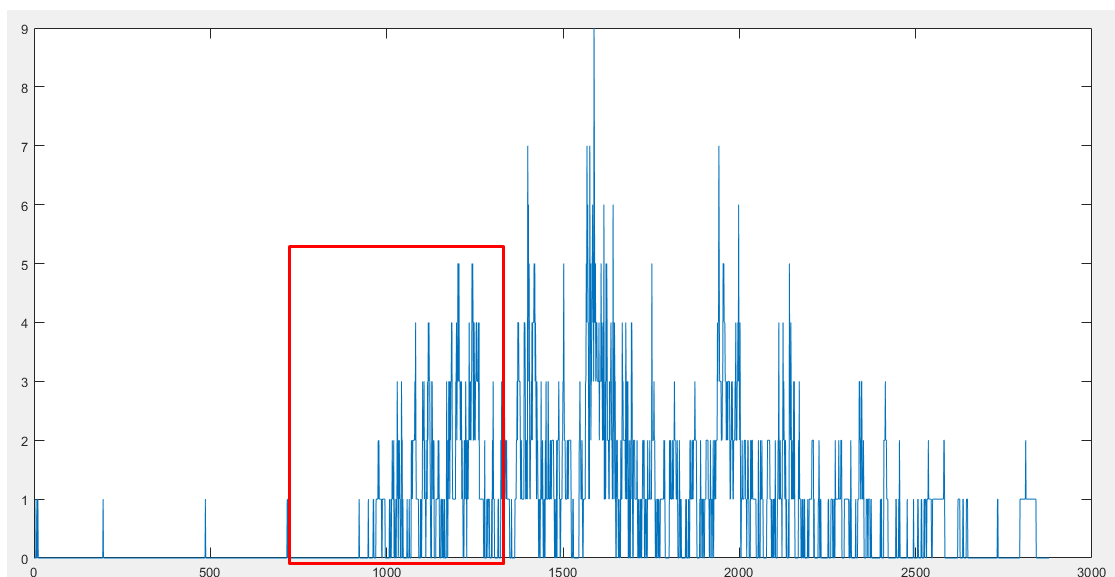


Рис 6.17 Реальні данні

В наступному експерименті значення $T_{\text{інф}}$ було фіксованим та вибрано значення 750. Після проведення аналогічних кроків було отримано наступні результати.

1. 74% відсотків спрогнозованих даних виявились вірними.
2. В 89% відсотках метод вдало запропонував змінити конфігурацію обладнання.
3. Час виконання алгоритму склав 37 секунд.

Результати показали, що метод з фіксованим $T_{\text{інф}}$ також має високі показники правдоподібності прогнозування та вчасної зміни конфігурації обладнання. Але час виконання програмної частини значно вищий від запропонованої варіації. Так як час саме в короткостроковому прогнозування є важливим критерієм, то метод з оптимальним $T_{\text{інф}}$ значно кращий для вчасної реакції на сплески навантаження.

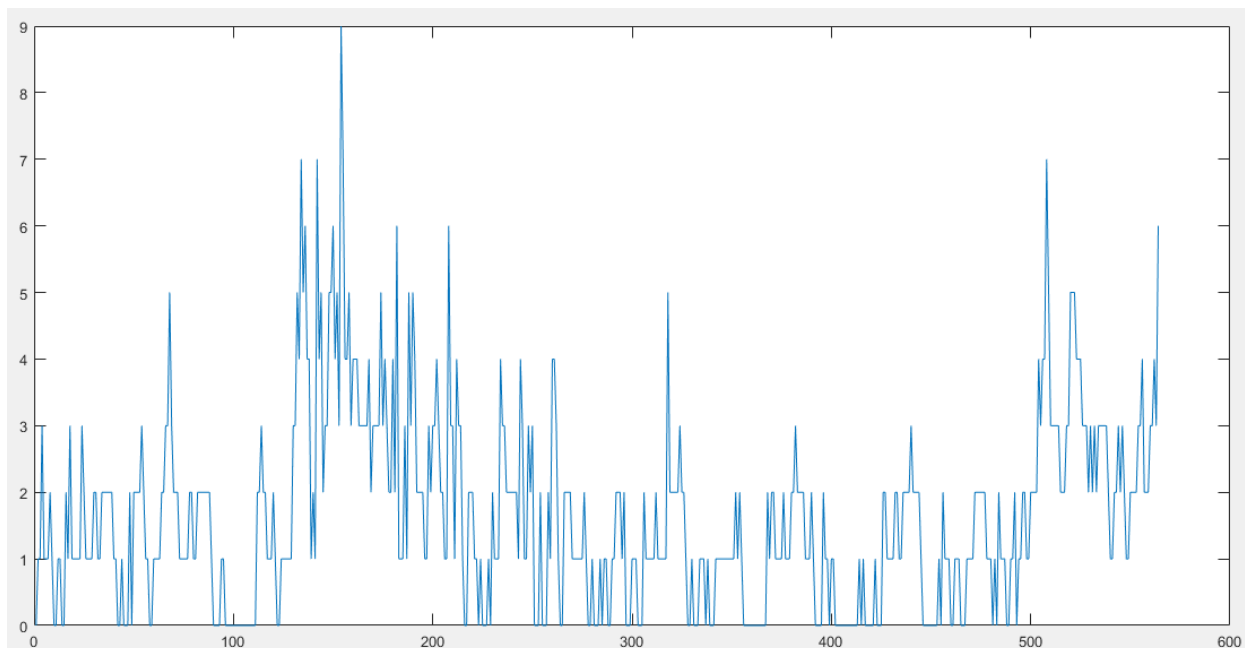


Рис 6.18 Результат роботи методу з фіксованим $T_{\text{інф}}$

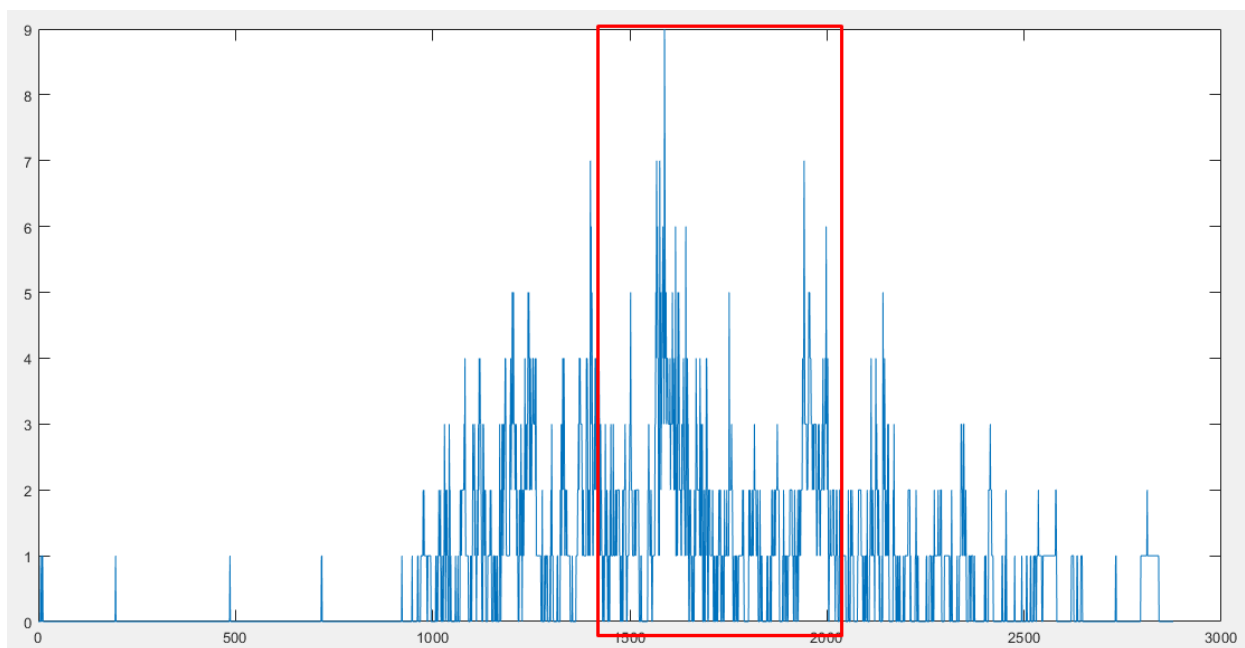


Рис 6.19 Реальні данні

Висновки

Проведено імітаційне моделювання запропонованих методів, виявлені обмеження запропонованих методів, які було враховано при створенні імітаційної моделі мережі обслуговуючих вузлів, в якій відтворені основні бізнес процеси обслуговування абонентів з використанням інфраструктури

NFV на основі статистичних даних реальної телекомунікаційної компанії. Якість системи оцінювалася за інтегральним показником згортки проф. А.М. Вороніна, та показала збільшення ефективності на 32%

Досліджено запропоновані архітектурні рішення щодо побудови телекомунікаційної мережі з віртуалізацією мережевих функцій та забезпеченням заданого рівня QoS, керуванням обчислювальними ресурсами ТКС, що показало доцільність проведення перерозподілу навантаження між групою фізичних та віртуальних вузлів обслуговування, покращення коефіцієнту використання обчислювальних ресурсів, а саме зменшення відсотку запитів, які було втрачено через перевищення допустимого часу обслуговування, а також зменшення відсотку часу низької завантаженості обчислювальних ресурсів.

Проведено математичне моделювання запропонованих процедур та алгоритмів, зокрема алгоритму планування обсягу ресурсів ГТС, процедур організації взаємодії підсистем ТКС, керування віртуалізованим телекомунікаційним середовищем із забезпеченням заданого рівня QoS та плануванням обсягу необхідних обчислювальних ресурсів для роботи ГТС. Результати моделювання запропонованих технічних рішень показали збільшення ймовірності утримання коефіцієнту використання ресурсів обслуговування у заданих межах.

ВИСНОВКИ

У дисертаційній роботі вирішено важливу науково-технічну проблему, пов'язану з розробкою наукових основ організації взаємодії елементів ТКС та віртуального зовнішнього обчислювального середовища, які базуються на систематизації методів керування гібридними телекомунікаційними сервісами та процесом їх обслуговування у ГТС для гарантування якісного обслуговування кінцевих користувачів та контрольованого використання обчислювальних ресурсів обслуговування.

За підсумками вирішення проблеми можна зробити наступні висновки:

1. Проведений аналіз стану ТКС та вимог, які висуваються користувачами телекомунікаційних послуг, показав відсутність підходів, гнучких моделей та методів контролю процесу обслуговування у ГТС із віртуалізацією мережевих функцій, де якість обслуговування користувачів залежить не лише від організації телекомунікаційної мережі, але й від організації обчислень, а також взаємодії між телекомунікаційним середовищем і обчислювальним середовищем.

2. Запропоновано методологію керування процесом обслуговування гібридних телекомунікаційних сервісів у ГТС, яка дозволяє організувати процес керування як фізичними ресурсами, так і хмарними сервісами та утворити середовище, яке забезпечує задані показники якості обслуговування.

3. Запропонована методологія забезпечення показників якості обслуговування гібридного сервісу у ГТС базується на чотирьох основних підходах: формування навантаження на вузол обслуговування, вибір потужності вузла обслуговування, визначення порядку роботи вузлів обслуговування та поточний контроль роботи системи. Це дозволило забезпечити якісне функціонування гетерогенного середовища обслуговування гібридних телекомунікаційних сервісів, а саме: зменшити

відсоток запитів, що обслуговуються в системі більше допустимого часу, визначеного експертами, - на 2%, зменшити відсоток запитів, які було втрачено через перевищення допустимого часу обслуговування на 3%, зменшити відсоток часу коли ресурси завантажені: менше за допустиме значення - на 8% та більше за допустиме значення на 10%.

4. Запропоновано модель керування інфраструктурою NFV, яка враховує особливості розташування дата центрів обслуговування, а також топологічної структури потоків гібридних телекомунікаційних сервісів, які поступають на обслуговування у дата центри, дозволяє забезпечити гнучке керування інформаційно-телекомунікаційною системою, організованою із застосуванням хмарних обчислень.

5. Розроблена архітектура локальної мережі оператора мобільного зв'язку дозволяє оцінити переваги та вузькі місця віртуалізації мережевих функцій, визначити особливості забезпечення показників якості обслуговування у ГТС.

6. Вдосконалено принцип керування телекомунікаційною системою, який на відміну від принципу програмно-керованих телекомунікаційних мереж розміщує у хмарному середовищі як всі функції керування службовим трафіком, так і процеси взаємодії між ними у відповідності до діючих стандартів; передаючи тільки результуючий керівний вплив на фізичні елементи, які здійснюють передавання інформаційних потоків. Всі підсистеми мобільного зв'язку є керованими із контролерів розташованих у дата центрі. Взаємодія між контролерами підсистем з метою керування відбувається лише в середині дата центру, що дозволить зменшити кількість службових потоків у телекомунікаційній мережі.

7. Запропоновано вперше математичну модель задачі формування вхідного потоку навантаження для використання ресурсів вузла обслуговування, яка дозволить розрахувати верхню межу для планування інтенсивності вхідного навантаження із урахуванням бажаної допустимої затримки заявок у системі обслуговування та використовувати

обчислювальні ресурси обслуговування у заданих межах. Обмеження навантаження відповідно до запропонованої моделі дозволило зменшити втрати гібридних ТК сервісів через перевищення черги на 5%.

8. Для вибору потужності обслуговуючого пристрою запропоновано вперше математичну модель, яка гарантує утримання процесу обслуговування на заданому рівні якості та вирізняється тим, що замість балансування навантаження між сталою кількістю обслуговуючих пристроїв враховує масштабованість віртуальних обчислювальних ресурсів вузлів обслуговування, мінімізуючи кількість цих пристроїв та втрати часу на балансування між обслуговуваними сутностями. Запропонована модель дозволяє розрахувати параметри системи обслуговування для будь якого вхідного потоку та забезпечити своєчасне надання сервісу.

9. Запропоновано принцип прогнозування необхідного віртуального ресурсу у хмарі для забезпечення роботи ГТС, який за рахунок використання даних системи моніторингу для розрахунку конфігурацій системи обслуговування та розкладу використання ресурсів з метою забезпечення показників якості, періодичного контролю достатності обраної конфігурації, забезпечить гнучкість та масштабованість системи керування процесом обслуговування гібридних телекомунікаційних сервісів у ГТС, а також використання віртуальних та фізичних обчислювальних ресурсів обслуговування у заданих межах.

10. Запропоновано вперше метод побудови розкладу залучення ресурсів на основі довгострокової статистики із забезпеченням показників якості обслуговування гібридних телекомунікаційних сервісів та використання ресурсів обслуговування у заданих межах, який за рахунок аналізу довгострокової статистики, ітераційної оцінки статистичних характеристик для інтервалів різної довжини, дозволяє сформувати розклад зміни конфігурації ресурсів обслуговування для вузла обслуговування у ГТС, та зменшити коефіцієнт простою ресурсів на 15%.

11. Вдосконалено авторегресійний метод із ковзковим математичним очікуванням, основна відмінність якого полягає у визначенні на основі довгострокової статистики інтервалу аналізу статистики, який забезпечує достовірність прогнозу із заданою ймовірністю, та мінімізує кількість точок аналізу, а відповідно й операцій при прогнозуванні у режимі реального часу, точність прогнозу, порівняно з фіксованим інтервалом аналізу, збільшилася на 3%, час розрахунку прогнозованого значення у режимі реального часу зменшився на 5%.

12. Запропоновано технологію керування ГТС, де обслуговування гібридних телекомунікаційних сервісів ведеться із застосуванням програмного забезпечення у багато хмарних дата центрах, технологія дозволила уникнути зниження якості обслуговування під час сплесків перевантаження та утримати показники якості обслуговування на заданому рівні за умови дотримання коефіцієнта використання ресурсів у заданих межах.

13. Досліджено запропоновані модифікації компонентів підсистеми PCRF (Policy and Charging Rules Function), нові процедури організації їх взаємодії між собою та з підсистемами керування віртуалізованим середовищем для забезпечення контролю якості обслуговування гібридних телекомунікаційних сервісів, проведено імітаційне та математичне моделювання процесів керування сервісами у ГТС, що дозволило підтвердити можливість запобігти погіршенню показників QoS на рівні абонентського пристрою кінцевого користувача та зменшити кількість відмов за рахунок скорочення часу реагування підсистеми керування телекомунікаційною та хмарною частинами гетерогенного середовища.

14. Досліджено запропоновані архітектурні рішення щодо побудови телекомунікаційної мережі з віртуалізацією мережевих функцій та забезпеченням заданого рівня QoS, керуванням обчислювальними ресурсами ТКС, що показало доцільність проведення перерозподілу навантаження між групою фізичних та віртуальних вузлів обслуговування, покращення

коефіцієнту використання обчислювальних ресурсів, а саме зменшення відсотку запитів, які було втрачено через перевищення допустимого часу обслуговування, а також зменшення відсотку часу низької завантаженості обчислювальних ресурсів.

15. Проведено математичне моделювання нових процедур та алгоритмів, зокрема алгоритму планування обсягу ресурсів ГТС, процедур організації взаємодії підсистем ТКС, керування віртуалізованим телекомунікаційним середовищем із забезпеченням заданого рівня QoS та плануванням обсягу необхідних обчислювальних ресурсів для роботи ГТС. Результати моделювання запропонованих технічних рішень показали збільшення ймовірності утримання коефіцієнту використання ресурсів обслуговування у заданих межах.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Рекомендации ITU-T M.3371 (10/2016) Требования к управлению услугами в системе управления электросвязью, совместимой с облаком. [Электронный ресурс]. – Режим доступа: <https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13064&lang=ru>.
2. Y.3500 : Information technology - Cloud computing - Overview and vocabulary. [Электронный ресурс]. – Режим доступа: <https://www.itu.int/rec/T-REC-Y.3500-201408-I/en>.
3. ETSI GS NFV 003: "Network Functions Virtualisation (NFV); Terminology for Main Concepts in. NFV".
4. Cloud computing framework for end to end resource management Recommendation ITU-T Y.3520. [Электронный ресурс]. – Режим доступа: <https://www.itu.int/rec/T-REC-Y.3520/en>.
5. ITU-T Y.2011 : General principles and general reference model for Next Generation Networks. [Электронный ресурс]. – Режим доступа: <https://www.itu.int/rec/T-REC-Y.2011-200410-I/en>.
6. ITU-T Y.2012 (04/2010) Функциональные требования и архитектура сетей последующих поколений. [Электронный ресурс]. – Режим доступа: <https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=10710&lang=ru>.
7. ITU-T Y.3502 Information technology - Cloud computing - Reference architecture. [Электронный ресурс]. – Режим доступа: <https://www.itu.int/rec/T-REC-Y.3502/en>.
8. ITU-T Y.3511 (03/14) : Framework of inter-cloud computing. [Электронный ресурс]. – Режим доступа: <https://www.itu.int/rec/T-REC-Y.3511/en>.
9. Pascal Kurtansky, Burkhard Stiller. State of the Art Prepaid Charging for IP Services // TIK-Report Nr. 236, November 2005.

10. H. Oumina and D. Ranc. Towards a Real Time Charging Framework for Complex Applications in 3GPP IP Multimedia Subsystem // Proceedings of the Conference on Next Generation Mobile Applications, Services and Technologies. September 2007. P. 145–150.

11. Gonzalo Camarillo, Miguel A. Garcia Martin. The 3G IP Multimedia subsystem (IMS): merging the Internet and cellular world.- 3rd ed. 2009. ISBN 978-0470-51662-1.

12. Khalid Al-Begain. IMS: development an deployment perspective, 1st edition. John Wiley& Sons, 2009 . P. 320.

13. Ільченко М.Ю. Сучасні телекомунікаційні системи./М.Ю. Ільченко, С.О. Кравчук – К.: НВП «Видавництво «Наукова Думка» НАН України. – 328 с.: іл.

14. Романов А.И. Основы теории телелкоммуникационных сетей: учебное пособие для вузов./ А.И. Романов– К.,2002 -152, ил. 84.

15. Основи керування мережами та послугами телекомунікацій: Підруч. для студ. вищ. навч. закл. за напрямком «Телекомунікації»/З ред. проф. Стеклова В.К.- К:Техніка, 2002.- 438 с.

16. Беркман Л.Н. Методи розрахунку показників якості конвергентних мереж на бази теорії ігор./ Л.Н. Беркман, О.М. Ткаченко, В.В. Григорович, П.Ю. Дещинський// Электроника и Связь. Тематический выпуска «Проблемы электроники» – ч.1. – 2008. – С. 220-221.

17. Вінницький В. П. Термінальне устаткування та передавання інформації в телекомунікаційних системах: Підручник./ В. П. Вінницький, В. Г. Поліщук – 2004. – 436 с.:аб.

18. Зайченко О.Ю. Аналіз та оптимізація показників якості та структур компютерних мереж з технологією АТМ: дисертація на здобуття наук.ступеня д-ра техн.наук: спец. 05.13.06/О.Ю. Зайченко. – К., 2005 р.

-
19. Hao Wang. 4G wireless video communication, 1st edition./ John Wiley and Sons Ltd , 2009. P. 320.
 20. Elwalid A. MATE: MPLS Adaptive Traffic Engineering / A. Elwalid, C. Jin, S. Low, I. Widjaja // IEEE INFOCOM – 2003.
 21. Урывский Л.А. Методика управления характеристиками обслуживания при изменении требований к качеству связи./ Л.А. Урывский, Е.А Прокопенко// Наукові записки УНДІЗ.– №3(11).– 2009.
 22. G.Priggouris, S.Hadjiefthymiades, LMarakos. Supporting IP QoS in the GPRS.// IEEE Network. – September 2009. –P.22-31.
 23. Elwalid A. MATE: MPLS Adaptive Traffic Engineering / A. Elwalid, C. Jin, S. Low, I. Widjaja // IEEE INFOCOM – 2003.
 24. Perkins C. IP Mobility Support for IPv4 / C. Perkins // RFC 3344. - August 2002. Checland P.B. Soft systems methodology: an overview // J. Appl. Syst. Anal. – 1988. –15. – P.27–36
 25. Fernando A. Hernández Solana. QoS and Radio Resource Management in Multimedia Packet Transmission for 3G Wireless IP Networks./ Fernando A. Hernández Solana, A. Valdovinos Bardaji, Casadevall Palacio.// IEEE VTC 2004 Spring. – Milán, Italia.
 26. Hemindez Solana, Antonio Valdovinos Bardajil, Femando Casadevall Palacio.Antennas and Propagation.// IEEE Network0-7803-8255-2/04/\$. – 2004. – Barcelona (Spain).
 27. Raymond Steele, Chin-Chun Lee, Peter Gould. GSM, cdmaOne and 3G Systems.// JOHN WILEY & SONS, LTD.– p. 498. ISBN 047149153
 28. ETSI 123.401. General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access.
 29. Friedhelm Hillebrand. GSM and UMTS: The Creation of Global Mobile Communication.// Wiley , 2002 - P. 371.

-
30. Generation Services using UMTS in a Converging Network Environment.//UMTS Forum, 2002 . – Report 14.
31. Maja Matijasevic, Lea Skorin-Kapov, Miran Mosmondor. Application-level QoS Negotiation and Signaling for Advanced Multimedia Services in the IMS.// IEEE Communications Magazine - July 2007, p. 25-33.
32. Plane Siddhartha Pandey, Vikram Jain, Debabrata Das Vincent Planat, Ragupathi Periannan. Performance Study of IMS Signaling.// IEEE Xplore Communications Magazine - July 2006, p. 123-135.
33. Основы передачи голосовых данных по сетям IP, 2-е изд.: Пер. с англ. – М.: ООО «И.Д. Вильямс», 2007.-400с.:ил. – Парал. тит. анагл. УДК. 681.3.07/ ISBN 978-5-8459-1281-7.
34. Руководство по технологиям объединенных сетей, 4-е издание.: Пер. с англ.: - М.: Издательский дом «Вильямс», 2005. – 1040 с.: ил.-Парал. тит. англ. ISBN 5-8459-0787-X.
35. M. Handley, H. Schulzrinne, E. Schooler, J. Rosenberg. SIP: Session Initiation Protocol.//RFC 2543, March 1999.
36. Tanir Ozcelebi, Igor Radovanovich. Multimedia Adaptation with SIP Resource Availability Signalling in IMS network.// IEEE 2007 International Conference on Convergence Information Technology. P. 1714-1719.
37. 3GPP Charging principles. [Электронный ресурс]: <http://www.scribd.com/doc/18148552/3GPP-Charging-Principles>.
38. H. Hakala, L. Mattila, J-P. Koskinen. Diameter Credit-Control Application.//RFC 4006, August 2005.
39. Robert Lloyd-Evans. QoS in intergrated 3G networks.//Artech House mobile communication series,2002. ISBN 1-58053-351-5.
40. Priggouris, G. Hadjiefthymiades, S. Merakos L. Supporting IP QoS in the GPRS.// IEEE Network September/October.Vol.14

-
41. Raymond Steele, Chin-Chun Lee, Peter Gould. GSM, cdmaOne and 3G Systems.// JOHN WILEY & SONS, LTD.– p. 498. ISBN 047149153
 42. 3GPP TS 32.299 R9. Telecommunication management.Charging management. Diameter Charging applicatio, P.143.
 43. ETSI 123.203 Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Policy and charging control architecture, p. 118
 44. 3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS); Tunnelling Protocol for Control plane (GTPv2-C); Stage 3 (Release 8).// 3GPP TS 29.274, 3rd Genral Partneship Project.P.143
 45. G.Priggouris, S.Hadjiefthymiades, LMarakos. Supporting IP QoS in the GPRS.// IEEE Network. – September 2009. –P.22-31.
 46. 3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS); Tunnelling Protocol for Control plane (GTPv2-C); Stage 3 (Release 8).// 3GPP TS 29.274, 3rd Genral Partneship Project.P.143.
 47. Sohel Q. Khan, Robert Gaglianello, Michael Luna. Experiences with Blending HTTP, RTSP, and IMS.//«IP Multimadia Systems (IMS) Infrastructure and Services» IEEE Communications Magazine. March 2007. P. 122-128.
 48. ETSI TS 136.331 LTE; Evolved Universal Terrestrial Radio Access (E-UTRA);Radio Resource Control (RRC).// Protocol specification: 3GPP TS 36.331 version 9.0.0 Release 9.
 49. Antonio Cuevas, Jose Ignacio Moreno, Hans Einsiedler. IMS Service Platform: A Solution for Next-Generation Network Operators to Be More than Bit Pipes.// IEEE Communication Magazine .2006. Aug. P.75-81.
 50. Syed A. Ahson, Mohammed Ilyas. IP Multimedia subsystem (IMS) handbook/ CRC Press, 2009. P. 250
 51. Глоба Л.С. Послуги в сучасних мобільних мережах України./ Л.С Глоба, О.О. Зінченко, О.М. Дяденко, І.М. Попова.// Електроника и связь.

Тематический выпуск «Электроника и нанотехнологии». – ч. 2. – 2009. – С. 291-296.

52. 3GPP TS 32.299 R9. Telecommunication management.Charging management. Diameter Charging applicatio, P.143.

53. Larysa Globa, Maria Skulish, Andrei Reverchuk. Server resources load monitoring serving different types of services// International Conference TCSET'2014 - February 2014.

54. Скулиш М.А. Організація роботи групи серверів для забезпечення потреб розподіленої системи тарифікації послуг // Наукові записки Українського науково-дослідного інституту зв'язку. – 2014. – №5(33), 120-126сс.

56. Добрушин Р.Л. О законе Пуассона для распределения частиц в пространстве / Добрушин Р.Л. // Укр.мат.ж. – 1956. – Т. 8. – С. 127–134 .

57. Breiman L. The Poisson tendency in traffic distribution / Breiman L. // Ann.Math. Statist. – 1963. – V.34, №1. – P. 308 – 311.

58. Клейнрок Л. Вычислительные системы с очередями. / Клейнрок Л. – М.: Мир. - 1979.

59. Клейнрок Л. Теория массового обслуживания. – М.: Машиностроение, 1979. – 432 с.

60. Бертсекас Д., Галлагер Р. Сети передачи данных. - М.: Мир, 1989.- 544с.

61. Форд Л., Фалкерсон Д. Потоки в сетях / Пер. с англ. – М.: Мир, 1966. – 276 с.

62. Гольдштейн А.Б., Гольдштейн Б.С. Технология и протоколы MPLS. - М.: Эко-Трендз, 2005. - 304 с.

63. Гольдштейн Б.С. Проблемы перехода к мультисервисным сетям // Вестник связи. – 2002. – № 12. – С 26-31.

64. Згуровский М.З. Микроволновые устройства телекоммуникационных систем / Згуровский М.З., Ильченко М.Е., Кравчук С.А. и др. – К.: ИВЦ «Видавництво “Політехніка”». - 2003. – Т.2: Устройства передающего и приемного трактов. Проектирование устройств и реализация систем. - 616с.

65. Ильченко М.Е. Сотовые радиосети в коммутацией пакетов. / Ильченко М.Е., Бунин С.Г., Войтер А.П. – К.: Наукова думка. – 2003, - 266с.

66. Ильченко М.Ю. Сучасні телекомунікаційні системи. / Ильченко М.Ю., Кравчук С.О. – К.: НВП «Винадвництво «Наукова думка» НАН України». – 328 с.

67. Бірюков Н.Л. Транспортне сети и системы электросвязи. Системы мультиплексирования: Ученик для студентів вузів по спеціальності «Телекомунікації» / Бірюков Н.Л., Стеклов В.К.; под ред. В.К. Стеклова. – К.; 2003, - 352 с.

68. Вінницький В.П. Термінальне устаткування та передавання інформації в телекомунікаційних системах : Підруч. для студ. вищ. навч. закл., які навчаються за напрямком «Телекомунікації». / Вінницький В.П., Поліщук В.Г. – К.: ИВЦ «Видавництво “Політехніка”». – 2004. – 436 с.

69. Дымарский Я.С., Крутякова Н.П., Яновский Г.Г. Управление сетями связи: принципы, протоколы, прикладные задачи. - М.: Эко-Трендз, 2003. - 384 с.

70. Дымарский Я. С. Проблемы оптимизации распределения ресурсов в сетях связи // Телекоммуникации. – 2002. – № 3. – С. 12-17.

71. Романов А.И. Управление потоками речевых сообщений на сетях связи. / Романов А.И. – К.:КВИУС. – 1998. – С. 272.

72. Романов А.И. Телекоммуникационные сети и управление. / Романов А.И. - Киевский университет. - 2003. – сс. 74-80.

73. Глоба Л.С. Математичні основи побудови інформаційно-телекомунікаційних систем / Посібник для студентів технічних спеціальностей/ Рек. МОН України, НТУУ “КПІ”, Інститут телекомунікацій, кафедра інформаційно-телекомунікаційних мереж, К.: Норіта-плюс, 2007. – 348 с.

74. Глоба Л.С. Концептуальное проектирование информационно-аналитических систем для сложных административных структур стратегического уровня управления: Науч. Монография. / Глоба Л.С., Голышев Л.К., Терновой М. Ю. – К. : ГП «Информационно аналитическое агенство», 2009. – 340с.

75. Урывский Л.А., Мошинская А.В. Синтез модели речевого сообщения в аксиоматике К.Шеннона // Зв’язок. – 2005. – № 6. – С. 47-48.

76. Кравчук С.О. Основы комп’ютерної техніки : Компоненти, системи, мережі : Навч. посіб. для студ. вищ. навч. зал. / Кравчук С.О., Шонін В.О. – К.: ІВЦ «Видавництво “Політехніка”» : «Видавництво “Каравела ”». – 2005. – 344с.

77. Концепція розвитку зв’язку України до 2010 року. – К.: Державний комітет зв’язку та інформатизації України. – 1999г. – 23 с.

78. Концепція конвергенції телефонних мереж і мереж з пакетною комутацією в Україні. – К.: Державний комітет зв’язку та інформатизації України. – 2003г. – 47 с.

79. Довгий С.А. Национальная информационная инфраструктура и развитие телекоммуникаций / Довгий С.А. // Вестник связи. – 2001. – № 10. – С 6-11.

80. Шварцман В.О. Интеграция в электросвязи. – М.: Агентство ИРИАС, 2001. – 167 с.

81. Концепція конвергенції телефонних мереж і мереж з пакетною комутацією в Україні. – К.: Державний комітет зв'язку та інформатизації України. - 2003. – 47 с.

82. Олифер В.Г. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. 2-е изд. / Олифер В.Г., Олифер Н.А. – СПб: Питер. - 2005. – 864 с.

83. Симоненко Д.В. Анализ многопутевых решений задач QoS-маршрутизации / Симоненко Д.В., Лемешко А.В., Дробот О.А. // Информационные технологии: наука, техника, технологии, образование, здоровье: Материалы XIV Международной научно-практической конференции (18-19 мая 2006 г.). – Харьков: Харьковский политехнический институт. - 2006. – С. 64.

84. Таненбаум Э. Компьютерные сети. / Таненбаум Э. - СПб.: Питер. – 2002. – 848 с.

85. Кучерявый Е.А. Управление трафиком и качество обслуживания в сети Интернет. / Кучерявый Е.А. – М.: Наука и Техника. – 2004. – 336 с.

86. Вегенша Ш. Качество обслуживания в сетях IP / Вегенша Ш.; пер. с англ.. - М.: Издательский дом «Вильямс». – 2003. - 386 с.

87. Дымарский Я.С. Оптимальное распределение ресурсов в сети с разнородными потоками. / Дымарский Я.С., Нурмиева М. В. // Вестник МАИСУ. – 2002. – №6.– С.31-35.

88. Засецкий А.В., Иванов А.Б., Постников С.Д., Соколов И.В. Контроль качества в телекоммуникациях и связи. Обслуживание, качество услуг, бизнес-управление. - М.: Syrus Systems, 2003. – 440 с.

89. Березко М.П. Математические модели исследования алгоритмов маршрутизации в сетях передачи данных / Березко М.П., Вишневский В.М., Левнер Е.В., Федотов Е.В. // Информационные процессы. – 2001. – Т.1, № 2. – С. 103–125.

-
90. Назаров А.Н. Модели и методы расчета структурно-сетевых параметров сетей АТМ. / Назаров А.Н. – М.: Горячая линия – Телеком. – 2002. – 256 с.
91. Назаров А.Н. АТМ: Принципы и технические решения создания сетей. / Назаров А.Н., Разживин И.А., Симонов М.В. – М.: Горячая линия – Телеком. – 2002. – 408 с.
92. ITU – T Recommendation Y.1540, IP Packet Transfer and Availability Performance parameters. December 2002.
93. ITU – T Recommendation Y.1541, Network Performance Objectives for IP – Based Service. May 2002.
94. Олифер В. Новые технологии и оборудование IP-сетей. / Олифер В., Олифер Н. - С.Пб.: BHV-СПб. – 2000. - 512 с.
95. Шварц М. Сети связи: протоколы, моделирование и анализ. Ч.1 / Шварц М.; пер. с англ. – М.: Наука, 1992. – 336с.
96. Дымарский Я.С. О некоторых задачах оптимизации АТМ-сетей / Дымарский Я.С., Нурмиева М. В. // Вестник МАИСУ.– 2002. – №8. – С. 20-31.
97. Харитонов С.А. Качество обслуживания и эффективное использование ресурсов в мультисервисных сетях / Харитонов С.А. // Вестник связи. – 2004. – №12. – С. 48-55.
98. J. Beran. Statistics for Long-Memory Processes / J. Beran. - New York, Chapman and Hall. – 1994.
99. O. J. Boxma. Regular Variation in a Multi-Source Fluid Queue / O. J. Boxma. - CWI Report BS-R9614. – 1996.
100. Estimators for long range dependence an empirical study by Murad S. Taqqu, Vadim Teverovsky and Walter Willinger in "Fractals". // Murad S. Taqqu, V. Teverovsky, W. Willinger. – Vol 3, No. 4. – 1995. – С. 785-788.

-
101. Maximum Likelihood estimation of stationary univariate fractionally integrated time series models by Fallaw Sowell. // Maximum Likelihood. – Journal of Econometrics 53. – 1992. – С. 165–188.
102. Лазарев В. Г. Динамическое управление потоками информации в сетях связи. / Лазарев В. Г., Лазарев Ю.В. – М.: Радио и связь. - 1983.
103. Б.В. Гнеденко. Лекции по теории массового обслуживания / Б.В. Гнеденко. – К. : КВИРТУ, 1960. – 206 с.
104. Grubber J. Delay Related Issues in Integration Voice and Data Networks: a review and some experimental work. / Grubber J. // In Proc. 6th Data Commun. Symp., Pacific Grove, CA, Nov. 27-29. – 1979. - P. 166-180.
105. Голд Б. Цифровые методы передачи речи / Голд Б. – ТИИЭР. – 1977. - Т.65. - №12. - С. 5-13.
106. Агаян А.А. Интегральные цифровые сети с коммутацией пакетов речи и данных. / Агаян А.А., Винник С.П., Родионов А.В. // Итоги науки и техники. Сер. Связь. – 1988. – Т. 2. – С. 109-149.
107. Ginsburg D. ATM solution for enternetworking. / Ginsburg D. - Addison Wesley Longman. – 1996. - P. 569.
108. Forgie J.W. Network speech: System implications of packetired speech // Forgie J.W., M.I.I. Lincoln Lab., Lexington, MA. – ANNU. Rep. ESD-TR-77-178 to the Defense Commun. Agency. - 1976.
109. Жданов А.Г. Передача речи по сетям с коммутацией пакетов (IP-телефония) / Жданов А.Г., Рассказов Д.А., Смирнов Д.А., Шипилов М.М.; Под ред. В.Ю.Бабкова, М.А.Вознюка – СПбГУТ. - СПб, 2001.
110. Рекомендация МСЭ-Т Н.320 (03/2004).
111. Феллер В. Введение в теорию вероятностей и ее приложения. / Феллер В. – Том 2. – М.: Мир. – 1967. – С. 751.
112. Б.С. Лившиц. Теория телектрафика. / Б.С. Лившиц, А.П.Пшеничников, А.Д. Харкевич. - М.: Связь, - 1979. – С. 360с.

-
113. Дымарский Я.С. Управление сетями связи: принципы, протоколы, прикладные задачи. / Дымарский Я.С., Крутякова Н.П., Яновский Г.Г. – М.: ИТЦ «Мобильные Коммуникации». – 2003. -384с.
114. Урывский Л.А., Мошинская А.В. Анализ зависимости дифференциальной энтропии дискретных отсчетов цифровизируемых сигналов от пространственно-энергетических параметров каналов связи // Збірка наукових праць ІПМЭ ім.Г.Є.Пухова. Моделювання та інформаційні технології. / Спеціальний випуск, № 2. – К.: ІПМЭ НАН України, 2005. – С.85-87.
115. Столингс В. Современные компьютерные сети. / Столингс В. – 2-е изд. – СПб.: Питер. – 2003. – 783 с.
116. Вишневский В.М. Теоретические основы проектирования компьютерных систем. / Вишневский В.М. – М.: Техносфера. – 2003. – 512 с.
117. Зарецкий К. А. Повышение качества функционирования сети связи при оптимальном распределении ресурсов. / Зарецкий К. А., Мейкшан В. И., Меленцова Н. А. // Электросвязь. – 2001. – № 3. – С. 34-38.
118. Б.С. Лившиц. Теория телетрафика. / Б.С. Лившиц, А.П. Пшеничников, А.Д. Харкевич. – М.: «Связь». – 1979г. – 163с.
119. R.Kalden. Wireless Access Based on GPRS / R.Kalden, I.Meirick, M.Meyer. - IEEE Personal Communications. [www.ee.mu.oz.au/pgrad/bpw/Data/Paper-GPRS.pdf] - 2001.
120. M.Zukerman. Circuit allocation and control in a hybrid switching system. / M.Zukerman. – Computer Network and ISDN System. - Vol.16, no. 4 – 1989. - PP. 281-298.
121. K.Chew. Performance analysis for GPRS with prioritized and non-prioritized mobility management procedures. / K.Chew, R.Tafazolli. - IEEE Personal Communications [www.ee.surrey.ac.uk/CCSR/Software/OPNET/chew-IEE-3G-2002.pdf]. - 2002.

-
122. C.H.Foh. Modeling and Performance Evaluation of GPRS / C.H.Foh, B.Meini, B.Wydrowski, M.Zukerman. - IEEE Personal Communications [http://netlab.caltech.edu/~bartek/Data/Paper-GPRS.pdf]. - 2001.
123. M.F. Neuts. Matrix-Geometric Solution in Stochastic Models: An algorithmic Approach / M.F. Neuts. - The Johns Hopkins University Press, Baltimore, MD. - 1981.
124. Гойхман В.Ю. Мультисервисные абонентские концентраторы. / Гойхман В.Ю., Гольдштейн А.Б. // Технологии и средства связи – 2004. - №1.
- A. A. Borovkov. Factorization identities and properties of the distribution of the supremum of sequential sums. / A. A. Borovkov. // Theor. Probability Appl. – 1970. – vol. 15. – pp. 359-402.
125. Vutukury S. A Simple Approximation to Minimum-Delay Routing / Vutukury S., Garcia-Luna-Aceves J.J. // Comput. Commun. Rev. – 1999. – Vol. 5, №. 1. – P. 227-238.
126. Vutukury S. MPATH: a loop-free multipath routing algorithm. / Vutukury S., Garcia-Luna-Aceves J.J. // Elsevier Journal of Microprocessors and Microsystems. – 2001. – № 24(6). – P. 319-327.
127. Vutukury S. SMART: a scalable multipath architecture for intra-domain QoS provisioning. / Vutukury S., Garcia-Luna-Aceves J.J. // Proc. of International Workshop on QoS in Multiservice IP Networks (QOS-IP 2001). – Rome. – 2001. – P. 67-79.
128. Wang Y. Explicit routing algorithms for Internet Traffic Engineering / Wang Y., Wang Z. // Proc. of 8th International Conference on Computer Communications and Networks. - Paris, 1999. - P. 582-588.
129. Younis O. Constraint-based routing in the internet: basic principles and recent research. / Younis O. Fahmy S. // IEEE Communication Society Surveys & Tutorials. - 2003. -Vol.5, №3. - P. 42-56.

-
130. Wang Z. Quality-of-service routing for supporting multimedia applications. / Wang Z., Crowcroft J. // IEEE JSAC. – 1996. – Vol. 14, № 7. – P. 1228-1234.
131. N. Likhanov. Analysis of an ATM buffer with self-similar ("Fractal") input traffic. / N. Likhanov, B. Tsybakov, N. D. Georganas. // Proc. IEEE INFOCOM'95, Boston, MA. – 1995. - pp. 985-992.
- A. G. Pakes, "On the tails of waiting-time distributions", J. Appl. Prob., vol. 12, pp. 555-564, 1975.
132. E. Willekens. Asymptotic expansions for waiting time probabilities in an M/G/1 queue with long-tailed service time / E. Willekens, J. L. Teugels // Queueing Systems. - vol. 10. – 1992. - pp. 295-312.
133. Авен О. И. Оценка качества и оптимизация вычислительных систем. / Авен О. И., Гурин Н. Н., Коган Я. А. – М.: Наука. – 1982. - 451 с.
134. Кондратенко А.И. Методические указания по курсу «Основы теории вычислительных систем и сетей ЭВМ». / Кондратенко А.И - Київ: «КПИ». – 1986. - 50 с.
135. Коффман Э. Г. Теория расписаний и вычислительные машины. / Коффман Э. Г. – М.: Наука. – 1984. - 334 с.
136. Цеглик Г. Г. Системы распределенных баз данных. / Цеглик Г. Г. - Львов: Свит. - 1990. - 168 с.
137. Демидович О.В. Моделювання управління маршрутизацією та навантаженням в глобальних інформаційно-обчислювальних мережах. / Демидович О.В., Цегелик Г.Т. // Праці 2-ї Укр. конф. з автоматичного керування "Автоматика-95". – Львів. -Т. 2. - 1995. - С.26.
138. Цегелик Г.Т. Построение математических моделей оптимального размещения копий файлов распределенных баз даннях / Цегелик Г.Т., Демидович А.В. // Автоматика и выч. техника. – Рига. - 1998. -№1.-С.53-63.

139. Демидович О.В. Критерій для визначення максимального числа копій кожного файла в розподіленій базі даних з копіями. / Демидович О.В., Цегелик Г.Г. // Вісн. НУ "Львівська політехніка". - Сер. інформаційні системи та мережі. - №406. - Львів. - 2000. - С.78-84.

140. Демидович О.В. Математичні моделі оптимального розподілу інформаційних ресурсів серед вузлів обчислювальних мереж та методи їх реалізації. / Демидович О.В. // Автореферат на здобуття наукового ступеня канд. техн. наук. - Львів. - 2001. - 20 с.

141. Theus M. Intrusion detection based on structural zeroes. / Theus M., Schonlau M. // Statistical Computing and Graphics Newsletter. - 1998. - №9(1). - С. 12-17.

142. Борзенко В.И. Многокритериальное управление локальной вычислительной сетью с использованием переменного управления к среде передачи. / Борзенко В.И., Косых В. С., Трахтенгерц Э. А., Шершаков В. И. // „Автоматика и телемеханика”, 1992, №7.

143. Larisa Globa, Mariia Slukysh. Nodal routing with traffic classification// Polish association for knowlage management Series: Studies&Proceedings №42, Bydgoszcz, Polska, - 2011, - pp. 37-46.

144. Суліма С. В. Гібридна система управління ресурсами для віртуалізованих мережевих функцій/ Суліма С. В., Скулиш М. А.// p-ISSN 1607-3274. Радіоелектроніка, інформатика, управління. - Запоріжжя -2017. - № 1 – сс.16-24.

145. Суліма С. В. Система управління ресурсами в центрах обробки даних оператора мережі мобільного зв'язку / Суліма С. В., Скулиш М. А.// Visnyk NTUU KPI Seriiia – Radiotekhnika Radioaparotobuduvannia, ISSN: 2310-0397, e-ISSN: 2310-0389, 2017, Iss. 68, pp. 27-32.

146. Application of the "Endless train" method for the SDN controller OpenDayLight M. Skulysh. // Visnyk NTUU KPI Seriiia – Radiotekhnika

Radioaparatobuduvannia, ISSN: 2310-0397, e-ISSN: 2310-0389, 2017, Iss. 69, pp. 32-38.

147. Скулиш. М.А. Заставенко А.А. Метод розподілу ресурсів сервера оператора мобільного зв'язку. // Вісник НТУУ "КПІ". Серія Радіотехніка, Радіоапаратобудування, ISSN: 2310-0397, e-ISSN: 2310-0389, № 60 (2015), сс..35-45.

148. M. Skulysh, L. Globa . Planning the loading of data centers' resources based on download statistics // Visnyk NTUU KPI Seriiia – Radiotekhnika Radioaparotobuduvannia, ISSN: 2310-0397, e-ISSN: 2310-0389, Iss 65 (2016), pp. 62-72.

149. Calculation functions of SDN controller for wireless backhaul infrastructure / Mariia Skulysh, Larysa Globa// Acta Electrotechnica et Informatica, Volume 17, Number 4, 2017, pp. 14-18(5), <https://doi.org/10.15546/aeei-2017-0029>.

150. Скулиш М. А. Організація роботи групи серверів для забезпечення потреб розподіленої системи тарифікації послуг. // Наукові записки Українського науково-дослідного інституту зв'язку. – 2014. – №5(33).- сс 56-64.

151. Скулиш М. А. Метод складання розкладу залучення ресурсів для високонавантажених інформаційних систем. // Наукові записки Українського науково-дослідного інституту зв'язку. – 2014. – №6(34). – сс.65-70.

152. Скулиш. М.А. Метод згладжування вхідного навантаження на сервер on-line тарифікації // Вісник Національного технічного університету «Львівська політехніка» Радіоелектроніка та телекомунікації, - Львів, - №796 , 2014, с. 97-103

-
153. Скулиш. М.А. Методи контролю перевантаженнями в процесі багатоетапної обробки заявок // Наукові записки Українського науково-дослідного інституту зв'язку. – 2015. – №1(35). – сс.81-86.
154. Скулиш М. А. Організація управління в багатоетапних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Вісник Національного технічного університету «Львівська політехніка» Радіoeлектроніка та телекомунікації — Львів, -2015. — № 818. — С. 80–85.
155. Skulysh M., Method of LTE functional units organization with evolved packet core virtualization. Skulysh M., Klimovych O. // Telecommunication Sciences. — 2015. — No. 4. — P. 38–45.
156. Skulysh M. Model for Efficient Allocation of Network Functions in Hybrid Environment / M. Skulysh, L. Globa, S. Sulima // Telecommunication Sciences -2016, - Volume 7, Number 1, сс. 39-45.
157. Скулиш М.А. Задача розподілу абонентського навантаження між базовими станціями з підтримкою SDR. / Скулиш М.А. Заставенко А.А. // Телекомунікаційні та інформаційні технології.– 2016. – №4, сс. 42-49.
158. Скулиш. М.А. Заставенко А.А. Метод контролю якості обробки інформаційних потоків у мережі 5G // Вісник Національного технічного університету «Львівська політехніка» Радіoeлектроніка та телекомунікації, - Львів, - №849, 2016 - с.265-273
159. Глоба Л.С. Система керування якістю обслуговування у складних гетерогенних телекомунікаційних системах/ Глоба Л.С. Вольвач Є. О., Скулиш М.А.// Системи управління, навігації та зв'язку. Збірник наукових праць. – Полтава: ПНТУ, 2017. – Т. 6 (46). – С. 158-162.
160. Формування зони обслуговування прийомо-передавальної станції в залежності від змінного вхідного навантаження. / М. А. Скулиш // Вісник Національного університету «Львівська політехніка». Радіoeлектроніка та телекомунікації. - Львів, - № 885 — 2017 – сс. . 50-54.

-
161. Hybrid resource provisioning system for telecommunication network / Skulysh M., Sulima S. // Сучасні інформаційні системи. – Харків -Том 2, № 1 - 2018 – СС. 47-51
162. Принцип обслуговування потоків у гетерогенному телекомунікаційному середовищі /Скулиш М.А., Романов О.І., Глоба Л.С. //ВЧЕНІ ЗАПИСКИ ТАВРІЙСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ ІМЕНІ В.І. ВЕРНАДСЬКОГО, Серія: Технічні науки, - Том 29 (68) № 2 - 2018 – сс. 92-98
163. Тимченко І.О. Концепції побудови сучасних мереж / Тимченко І.О., Скулиш М.А., Глоба Л.С. // Радиоэлектроника и информатика = Radioelectronics & informatics : науч. журн. / Харьк. нац. ун-т радиоэлектроники. – Харків: ХНТУРЭ, 2018. - №1, - сс. 25-29
164. Принцип резервування ресурсів у віртуалізованому середовищі для контролю показників обслуговування / М.А. Скулиш, О.І. Романов, Л.С. Глоба // Вісник Інженерної академії України. – 2018 – сс. 197-202
165. Шилов Ф. І. Дослідження ефективності методу оптимального вибору обчислювальних ресурсів для білінгових систем / Ф. І. Шилов, М. А. Скулиш, А. Сафарян // Системи управління, навігації та зв'язку. - 2018. - Вип. 3. - С. 147-152.
166. Математична модель пошуку оптимального обсягу ресурсів віртуального телекомунікаційного вузла обслуговування/ М.А. Скулиш // Сучасні інформаційні системи. – Харків -Том 2, № 2- 2018 – сс. 30-34.
167. Скулиш М.А. Романов О.І., Нестеренко М.М. Принцип прогнозування необхідного віртуального ресурсу хмарної системи для оператора мобільного зв'язку // Збірник наукових праць ВІПІ. – Київ, 2018. – Вип. №2 – С. 113 – 119.
168. Математичні методи аналізу та керування телекомунікаційними мережами. Л.С. Глоба, О.М. Дяденко, А.Ю. Пилипенко, М.А. Скулиш. -

Видавництво Інститут обдарованої дитини НАПН України - ISBN 978-966-2633-84-9 Монографія (365).

169. Система обробки мультисервісних інформаційних потоків. Патент на корисну модель №85488. Патент опубліковано 25.11.2013, бюл. № 22/2013 Скулиш М.А. Глоба Л.С. Ватула А.С, Марчук А.П.

170. Спосіб зваженого кругового обслуговування черг мультисервісних інформаційних потоків патент Патент на корисну модель №85487. Патент опубліковано 25.11.2013, бюл. № 22/2013.

Скулиш М.А. Глоба Л.С. Ватула А.С, Марчук А.П.

171. Глоба Л. С., Скулиш М. А. Керування трафіком в мультисервісному комутаційному центрі //Telecommunication Sciences. – 2011. – Т. 2. – №. 1. – С. 30.

172. Скулиш М. А. Метод управління ресурсами в датацентрах оператора мережі мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Radioelectronics & Informatics Journal (ISSN 1563-0064), - Харьков — 2015. — № 3. — С. 8-13

173. Скулиш. М.А. Заставенко А.А. Пилипенко А.Ю. Метод выбора контейнера для миграции виртуальной машины в облачном хранилище данных.// Вісник Харківського національного університету імені В.Н. Каразіна,

серія "Математичне моделювання. Інформаційні технології. Автоматизовані системи управління" випуск 31 - 2016 р., - с.47-58

174. Метод організації функцій контролеру SDN «Нескінчений потяг» Скулиш М. А. - Radioelectronics & Informatics Journal (ISSN 1563-0064), - Харьков — 2016. — № 3. — С. 21–28.

175. L. S. Globa. Paralle computing usage for computation of the queues QoS parameters / L. S. Globa ; K. O. Yermakova ; V. V. Kushnyr ; M. A. Skulysh

// 20th International Crimean Conference "Microwave & Telecommunication Technology" – 2010 –p. 466 – 467 DOI: 10.1109/CRMICO.2010.5632345

176. L. S. Globa. Efficiency increase for the choice of the virtual connection direction // L. S. Globa ; M. A. Skulysh // 20th International Crimean Conference "Microwave & Telecommunication Technology" – 2010 – p. 502 - 503 DOI: 10.1109/CRMICO.2010.5632474

177. L. S. Globa. Improving efficiency service of queue in nodal centers / L. S. Globa ; M. A. Skulysh // 21st International Crimean Conference "Microwave & Telecommunication Technology" - 2011– p. 461 – 462

178. L. S. Globa. Estimation scheme for improvement of weighted round robin algorithm for requests service in the queues / L. S. Globa ; M. A. Skulysh // 22nd International Crimean Conference "Microwave & Telecommunication Technology" - 2012– p.393 – 394

179. L.S. Globa. Method of the analysis of power consumption dependence from service of a number of processes / L.S. Globa ; M.A. Skulysh // 23rd International Crimean Conference "Microwave & Telecommunication Technology" - 2013– p.432 – 433

180. L.S. Globa. Method of increase of fault tolerance of mobile communication systems at the moments of peak load / L.S. Globa ; M.A. Skulysh ; A.V. Reverchuk // 23rd International Crimean Conference "Microwave & Telecommunication Technology" – 2013 – p.533 – 534.

181. M.A. Skulysh. Method of increase of efficiency of assistance to the victims according to the integrated number 112 // 23rd International Crimean Conference "Microwave & Telecommunication Technology" – 2013 – p. 545 – 546

182. Globa Larisa. Control strategy of the input stream on the online charging system in peak load moments / Globa Larisa ; Skulysh Mariia ;

Reverchuk Andriy // 24th International Crimean Conference Microwave & Telecommunication Technology – 2014 – p. 312 – 313

183. Larysa G. et al. Managing of incoming stream applications in online charging system / Globa Larysa ; Skulysh Mariia ; Pidgurska Tetiana ; Andrii Reverchuk //Telecommunications (BIHTEL), 2014 X International Symposium on. – IEEE, 2014. – C. 1-6.

184. Skulysh M. Management of Multiple Stage Queuing Systems / M. Skulysh, S. Sulima // CADSM 2015 : 13-th International conference, 24–27 February 2015 : conference proceedings. — Lviv–Polyana, 2015. — P. 431– 434.

185. Globa Larisa. The method of resources allocation for processing requests in online charging system / Globa Larisa, Skulysh Mariia, Zastavenko Anna// CADSM 2015 : 13-th International conference, 24–27 February 2015 : conference proceedings. — Lviv–Polyana, 2015. — P. 211 – 213

186. Skulysh M., Klimovych O. Approach to virtualization of Evolved Packet Core Network Functions //Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), 2015 13th International Conference The. – IEEE, 2015. – C. 193-195.

187. Mariia S., Svitlana S. Service deployment aspects in the systems with network function virtualization //Radio Electronics & Info Communications (UkrMiCo), 2016 International Conference. – IEEE, 2016. – C. 1-7.

188. Globa L. Method for resource allocation of virtualized network functions in hybrid environment / L. Globa, M. Skulysh, S. Sulima // 2016 IEEE International Black Sea Conference on Communications and Networking Pages: 1 - 5, DOI: 10.1109/BlackSeaCom.2016.7901546.

189. Skulysh M. The Method of Computing Organization in High Loaded SDN Controller System / // CADSM 2017 : 13-th International conference, 24–27 February 2017 : conference proceedings. — Lviv–Polyana, 2017. — P. 431– 434.

190. Mariia Skulysh. The method of resources involvement scheduling based on the long-term statistics ensuring quality and performance parameters //Radio Electronics & Info Communications (UkrMiCo), 2017 International Conference. – С.

191. Konstantin Sunduchkov, Sampling for direct search method of all system implementations developed according to the requirements of the technical design specification / Konstantin Sunduchkov ; Mariia Skulysh ; Svitlana Sulima ; B. Savchuk // 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T) - 2017 - P: 83 – 86

192. Skulysh Mariia. The structure of a mobile provider network with network functions virtualization / Skulysh Mariia ; Romanov Oleksandr // 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET) – 2018 – P: 1032 - 1034

193. Скулиш М. А. Організація управління в багатоетапних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Проблеми телекомунікацій ПТ-2015 : 9-а міжнародна науково-технічна конференція, — Київ, 2015. — С. 319–321.

194. Скулиш М. А. Управління ресурсами в віртуалізованих мережах оператора мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Інфокомунікації – сучасність та майбутнє : 5-а міжнародна науково-практична конференція, 29-30 жовтня 2015 : матеріали конференції. — Одеса, 2015. — С.84–88.

195. Суліма С. В. Алгоритм відображення та планування віртуалізованих функцій в мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Проблеми телекомунікацій ПТ-2016 : 10-а міжнародна науково-технічна конференція, 19–22 квітня 2016 : матеріали конференції. — Київ, 2016. — С. 372–374.

196. Skulysh M. Service deployment aspects in the systems with network function virtualization / M. Skulysh, S. Sulima //

RadioElectronics&InfoCommunications : First International Conference, 11–16 September 2016 : conferenceproceedings. — Kyiv, 2016. — P. 1–7.

197. Скулиш М.А, Оцінка залежності використання ресурсів від вхідного навантаження на сервері мобільного оператора/ Скулиш М.А, Заставенко А.А., Яндович І.Є. // Проблеми телекомунікацій ПТ-2014 : 8-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2014 . — С. 202–206.

198. Скулиш М.А, Схема згладжування вхідного навантаження на сервер мобільного оператора / Глоба Л.С., Скулиш М.А., Реверчук А.В. // Проблеми телекомунікацій ПТ-2014 : 8-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2014 . — С. 46-49.

199. Тимченко І.О. Метод відновлення вузла у мережах NFV / Тимченко І.О., Скулиш М.А. // Проблеми телекомунікацій ПТ-2017 : 11-а міжнародна науково-технічна конференція, 18–21квітня 2017 : матеріали конференції. — Київ, 2017 . — С. 262–264.

200. Скулиш М.А, Анализ методов прогнозирования с использованием нейронной сети/ Запорожец Д.Б., Скулиш М.А. // Проблеми телекомунікацій ПТ-2018 : 10-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2018 . — С. 266-268.

201. Скулиш М.А, Впровадження віртуалізації мережевих функцій на мережі платформи передачі SMS – повідомлень / Чиж В.В. , Скулиш М.А.// Проблеми телекомунікацій ПТ-2018 : 10-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2018 . — С. 272-275.

202. Скулиш М.А, Планування ресурсів гетерогенного телекомунікаційного середовища на базі аналізу статистичних даних / Белокур Б.В., Скулиш М.А.// Проблеми телекомунікацій ПТ-2018 : 10-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2018 . — С. 279-281.

203. Скулиш М.А, Сравнительный анализ методов краткосрочного прогнозирования сетевого трафика / Запорожец Д.Б., Скулиш М.А. // Проблемы телекоммуникаций ПТ-2018 : 10-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2018 . — С. 310-312.

204. Скулиш М.А, Аналіз засобів моделювання для мережевої СМО із динамічно змінюваними параметрами вузлів обслуговування/ Нездвецький В.С., Скулиш М.А. // Проблемы телекоммуникаций ПТ-2018 : 10-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2018 . — С. 310-312.

205. Скулиш М.А, Проблемы впровадження системи екстреної допомоги населенню за єдиним номером 112 // Проблемы телекоммуникаций ПТ-2013 : 7-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2013 . — С. 138-140.

206. Скулиш М.А. Система комплексного керування інформаційними потоками в комутаційному центрі / Глоба Л.С., Скулиш М.А. // Проблемы телекоммуникаций ПТ-2013 : 7-а міжнародна науково-технічна конференція, матеріали конференції. — Київ, 2013 . — С. 141-144.

207. Скулиш М.А. Существующие основные подходы по методам оценки качества передачи речевой информации в сетях IP / Гайдаенко О.В., Скулиш М.А. // IV Міжнародна конференція молодих вчених «Електроніка-2013». Збірник статей. – 2013 - 302-305сс.

208. Скулиш М.А. Математична модель методу оптимізації обслуговування викликів у «Системі 112»/ Марчук А.П, Скулиш М.А. // IV Міжнародна конференція молодих вчених «Електроніка-2013». Збірник статей. – 2013 - 302-305сс.

209. Жожикашвили В.А. Сети массового обслуживания. Теория и применение к сетям ЭВМ. / Жожикашвили В.А., Вишневский В. М. - М., “Радио и связь”. – 1988.

-
210. Подиновский В. В. Парето-оптимальные решения многокритериальных задач. / Подиновский В. В., Ногин В. Д. – М., «Наука», 1982.
211. Михайлов В.Г. Измерение параметров речи // Михайлов В.Г., Златоустова Л.В.; под ред. М.А.Сапожкова. – М.: Радио и связь. – 1987. – 167с.
212. Максимов В.В. Імітаційне моделювання електронних компонентів, вузлів та пристроїв в середовищі пакету Electronics Workbench / Максимов В.В., Вульте О.А.; навчально-методичний посібник. – К.; ІВЦ «Політехніка», 2008. – 108с.
213. Хмелёв К.Ф. Основы SDN: Монографія. / Хмелёв К.Ф. – К.: ІВЦ «Видавництво “Політехніка”». – 2003. – 584с.
214. Максимов В.В. Цифрові електронні комутаційні системи MT20/25, EWSD, ALCATEL 1000 E-10, 5ESS. Приклади розрахунку навантаження на обладнання: Навч.-метод. посіб. / Максимов В.В. – К.: ІВЦ «Видавництво “Політехніка”». – 2004. – 72с.
215. ДЕМИДОВ, ІВАН ВАСИЛЬОВИЧ. Синтез телекомунікаційних сервісних платформ національного масштабу. Дисертація на здобуття доктора технічних наук. 2017 - lp.edu.ua
216. ЛАВРІВ, Орест Андрійович. Методи та моделі надання послуг в гетерогенних розподілених інформаційно-телекомунікаційних системах. Дисертація на здобуття доктора технічних наук, - 2018 - ena.lp.edu.ua
217. КИРИК, Мар'ян Іванович. Методи та моделі управління контентом в розподілених інфокомунікаційних системах, Дисертація на здобуття доктора технічних наук, - 2018 - ena.lp.edu.ua
218. 3GPP Specification TS: 23.107 Quality of Service (QoS) concept and architecture

<https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=783>

219. 3GPP Specification TS: 23.402 Architecture enhancements for non-3GPP accesses

<https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=850>

ДОДАТОК 1

Частина 1. Код імітаційної моделі GPSS

* СМО

Stor1 STORAGE 10000; оголошення багатоканального пристрою з 10000 каналами
(ресурсу 1 типу об'ємом 10000)

Stor2 STORAGE 10000; оголошення багатоканального пристрою з 10000 каналами
(ресурсу 2 типу об'ємом 10000)

Stor3 STORAGE 10000; оголошення багатоканального пристрою з 10000 каналами
(ресурсу 3 типу об'ємом 10000)

Stor4 STORAGE 10000; оголошення багатоканального пристрою з 10000 каналами
(ресурсу 4 типу об'ємом 10000)

StorS1 STORAGE 10000; оголошення багатоканального пристрою з 10000
каналами (ресурсу 1 типу об'ємом 10000)

StorS2 STORAGE 10000; оголошення багатоканального пристрою з 10000
каналами (ресурсу 2 типу об'ємом 10000)

StorS3 STORAGE 10000; оголошення багатоканального пристрою з 10000
каналами (ресурсу 3 типу об'ємом 10000)

StorS4 STORAGE 10000; оголошення багатоканального пристрою з 10000
каналами (ресурсу 4 типу об'ємом 10000)

***Параметри обслуговування в залежності від типу заявки (4 типи)

*1 блок

Pr_time1 FUNCTION P1,L4; час обслуговування на блоці 1 в залежності від типу
заявки (який записаний у першому параметрі транзакту)

1,0.002/2,0.002/3,0.002/4,0.002

Vol1_1 FUNCTION P1,L4; кількість необхідного ресурсу типу 1 на блоці 1 в
залежності від типу заявки (який записаний у першому параметрі транзакту)

1,1000/2,1000/3,1000/4,1000

Vol2_1 FUNCTION P1,L4; кількість необхідного ресурсу типу 2 на блоці 1 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,5/2,5/3,5/4,5

Vol3_1 FUNCTION P1,L4; кількість необхідного ресурсу типу 3 на блоці 1 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,1000/2,1000/3,1000/4,1000

Vol4_1 FUNCTION P1,L4; кількість необхідного ресурсу типу 4 на блоці 1 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0/2,0/3,0/4,0

*2 блок

Pr_time2 FUNCTION P1,L4; час обслуговування на блоці 2 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0/2,0.3/3,3.6/4,0.001

Vol1_2 FUNCTION P1,L4; кількість необхідного ресурсу типу 1 на блоці 2 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0/2,0/3,0/4,0

Vol2_2 FUNCTION P1,L4; кількість необхідного ресурсу типу 2 на блоці 2 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0/2,5/3,6/4,3

Vol3_2 FUNCTION P1,L4; кількість необхідного ресурсу типу 3 на блоці 2 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0/2,1000/3,2000/4,500

Vol4_2 FUNCTION P1,L4; кількість необхідного ресурсу типу 4 на блоці 2 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0/2,1000/3,2000/4,500

*3 блок

Pr_time3 FUNCTION P1,L4; час обслуговування на блоці 3 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,1/2,1/3,1/4,1

Vol1_3 FUNCTION P1,L4; кількість необхідного ресурсу типу 1 на блоці 3 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,500/2,1000/3,1000/4,600

Vol2_3 FUNCTION P1,L4; кількість необхідного ресурсу типу 2 на блоці 3 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,100/2,200/3,500/4,200

Vol3_3 FUNCTION P1,L4; кількість необхідного ресурсу типу 3 на блоці 3 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,500/2,1000/3,1000/4,600

Vol4_3 FUNCTION P1,L4; кількість необхідного ресурсу типу 4 на блоці 3 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0/2,0/3,0/4,0

*4 блок

Pr_time4 FUNCTION P1,L4; час обслуговування на блоці 4 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0.5/2,0.5/3,0.5/4,0.5

Vol1_4 FUNCTION P1,L4; кількість необхідного ресурсу типу 1 на блоці 4 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,1000/2,1000/3,1000/4,1000

Vol2_4 FUNCTION P1,L4; кількість необхідного ресурсу типу 2 на блоці 4 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,100/2,10/3,100/4,100

Vol3_4 FUNCTION P1,L4; кількість необхідного ресурсу типу 3 на блоці 4 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,100/2,100/3,100/4,100

Vol4_4 FUNCTION P1,L4; кількість необхідного ресурсу типу 4 на блоці 4 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,100/2,100/3,100/4,100

*5 блок

Pr_time5 FUNCTION P1,L4; час обслуговування на блоці 5 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0.2/2,0.2/3,0.2/4,0.2

Vol1_5 FUNCTION P1,L4; кількість необхідного ресурсу типу 1 на блоці 5 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,200/2,1000/3,2000/4,400

Vol2_5 FUNCTION P1,L4; кількість необхідного ресурсу типу 2 на блоці 5 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,100/2,200/3,600/4,150

Vol3_5 FUNCTION P1,L4; кількість необхідного ресурсу типу 3 на блоці 5 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,200/2,1000/3,2000/4,400

Vol4_5 FUNCTION P1,L4; кількість необхідного ресурсу типу 4 на блоці 5 в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,0/2,0/3,0/4,0

MaxDel FUNCTION P1,D4; максимальний допустимий час обслуговування в залежності від типу заявки (який записаний у першому параметрі транзакту)

1,20/2,5/3,10/4,20

INCLUDE "dataParam.txt"; зчитування додаткових комад з файлу dataParam.txt (визначення функцій параметрів розподілу заявок в залежності від години)

CheckS1 BVARIABLE

((R\$Stor1'GE'FN\$Vol1_1)'AND'(R\$Stor2'GE'FN\$Vol2_1)'AND'(R\$Stor3'GE'FN\$Vol3_1)'AND'(R\$Stor4'GE'FN\$Vol4_1)); якщо

Stor1>=Vol1_1,Stor2>=Vol2_1,Stor3>=Vol3_1,Stor4>=Vol4_1 то істина

CheckS2 BVARIABLE

((R\$Stor1'GE'FN\$Vol1_2)'AND'(R\$Stor2'GE'FN\$Vol2_2)'AND'(R\$Stor3'GE'FN\$Vol3_2)'AND'(R\$Stor4'GE'FN\$Vol4_2)); якщо

Stor1>=Vol1_2,Stor2>=Vol2_2,Stor3>=Vol3_2,Stor4>=Vol4_2 то істина

CheckS3 BVARIABLE

((R\$Stor1'GE'FN\$Vol1_3)'AND'(R\$Stor2'GE'FN\$Vol2_3)'AND'(R\$Stor3'GE'FN\$Vol3_3)'AND'(R\$Stor4'GE'FN\$Vol4_3)); якщо

Stor1>=Vol1_3,Stor2>=Vol2_3,Stor3>=Vol3_3,Stor4>=Vol4_3 то істина

CheckS4 BVARIABLE

((R\$Stor1'GE'FN\$Vol1_4)'AND'(R\$Stor2'GE'FN\$Vol2_4)'AND'(R\$Stor3'GE'FN\$Vol3_4)'AND'(R\$Stor4'GE'FN\$Vol4_4)); якщо

Stor1>=Vol1_4,Stor2>=Vol2_4,Stor3>=Vol3_4,Stor4>=Vol4_4 то істина

CheckS5 BVARIABLE

((R\$Stor1'GE'FN\$Vol1_5)'AND'(R\$Stor2'GE'FN\$Vol2_5)'AND'(R\$Stor3'GE'FN\$Vol3_5)'AND'(R\$Stor4'GE'FN\$Vol4_5)); якщо

Stor1>=Vol1_5,Stor2>=Vol2_5,Stor3>=Vol3_5,Stor4>=Vol4_5 то істина

CheckSS1 BVARIABLE

((R\$StorS1'GE'FN\$Vol1_1)'AND'(R\$StorS2'GE'FN\$Vol2_1)'AND'(R\$StorS3'GE'FN\$Vol3_1)'AND'(R\$StorS4'GE'FN\$Vol4_1)); якщо

StorS1>=Vol1_1,StorS2>=Vol2_1,StorS3>=Vol3_1,StorS4>=Vol4_1 то істина

CheckSS2 BVARIABLE

((R\$StorS1'GE'FN\$Vol1_2)'AND'(R\$StorS2'GE'FN\$Vol2_2)'AND'(R\$StorS3'GE'FN\$Vol3_2)'AND'(R\$StorS4'GE'FN\$Vol4_2)); якщо

StorS1>=Vol1_2,StorS2>=Vol2_2,StorS3>=Vol3_2,StorS4>=Vol4_2 то істина

CheckSS3 BVARIABLE

((R\$StorS1'GE'FN\$Vol1_3)'AND'(R\$StorS2'GE'FN\$Vol2_3)'AND'(R\$StorS3'GE'FN\$Vol3_3)'AND'(R\$StorS4'GE'FN\$Vol4_3)); якщо

3_3)'AND'(R\$StorS4'GE'FN\$Vol4_3)); якщо

StorS1>=Vol1_3,StorS2>=Vol2_3,StorS3>=Vol3_3,StorS4>=Vol4_3 то істина

CheckSS4 B VARIABLE

((R\$StorS1'GE'FN\$Vol1_4)'AND'(R\$StorS2'GE'FN\$Vol2_4)'AND'(R\$StorS3'GE'FN\$Vol3_4)'AND'(R\$StorS4'GE'FN\$Vol4_4)); якщо

StorS1>=Vol1_4,StorS2>=Vol2_4,StorS3>=Vol3_4,StorS4>=Vol4_4 то істина

CheckSS5 B VARIABLE

((R\$StorS1'GE'FN\$Vol1_5)'AND'(R\$StorS2'GE'FN\$Vol2_5)'AND'(R\$StorS3'GE'FN\$Vol3_5)'AND'(R\$StorS4'GE'FN\$Vol4_5)); якщо

StorS1>=Vol1_5,StorS2>=Vol2_5,StorS3>=Vol3_5,StorS4>=Vol4_5 то істина

Ttime TABLE M1,0,1,70; таблиця часу перебування транзакта в системі(70 одиничних інтервалів)

Birth TABLE AC1,0,300,2016; таблиця абсолютного часу генерації транзакта в системі(2016 інтервалів величиною в 5 хвилин(300 секунд))

GENERATE (Exponential(1,0,(FN\$Lyambda1#FN\$LyambD1))); генерація потоку заявок по експоненційному закону без зсуву з мат.очікуванням, що розраховується функцією Lyambda1 та помножений на коефіцієнт LyambD1

ASSIGN 1,1; запис в параметр P1 типу сервісу

TRANSFER ,Tab_birth; перехід на Tab_birth

GENERATE (Exponential(1,0,(FN\$Lyambda2#FN\$LyambD2))); генерація потоку заявок по експоненційному закону без зсуву з мат.очікуванням, що розраховується функцією Lyambda2 та помножений на коефіцієнт LyambD2

ASSIGN 1,2; запис в параметр P1 типу сервісу

TRANSFER ,Tab_birth; перехід на Tab_birth

GENERATE (Exponential(1,0,(FN\$Lyambda3#FN\$LyambD3))); генерація потоку заявок по експоненційному закону без зсуву з мат.очікуванням, що розраховується функцією Lyambda3 та помножений на коефіцієнт LyambD3

ASSIGN 1,3; запис в параметр P1 типу сервісу

TRANSFER ,Tab_birth; перехід на Tab_birth

GENERATE (Exponential(1,0,(FN\$Lyambda4#FN\$LyambD4))); генерація потоку заявок по експоненційному закону без зсуву з мат.очікуванням, що розраховується функцією Lyambda4 та помножений на коефіцієнт LyambD4

ASSIGN 1,4; запис в параметр P1 типу сервісу

* TRANSFER ,Tab_birth; перехід на Tab_birth

*P1 - параметр транзакта, що відображає тип сервісу

Tab_birth TABULATE Birth; збір статистики

*****Блок

1*****

*met1 QUEUE wait1; зайняття черги

met1 SEIZE Channel1; зайняття приладу Channel1

*AvTest TEST E FN\$OnOff,1,met_Test1; якщо значення OnOff 1, то йдемо далі, інакше - на met_Test1

* TEST E BV\$CheckS1,1,metS1; якщо значення булевої змінної 1, то йдемо далі, інакше - на metS1

met_Test1 TEST E BV\$CheckS1,1; якщо значення булевої змінної 1, то йдемо далі, інакше - очікування

TEST GE FN\$MaxDel,M1,discard1; якщо час перебування в системі не перевищує допустимого значення, то йдемо далі, інакше - на мітку discard1

ENTER Stor1,FN\$Vol1_1; зайняття Vol1_1 ресурсів багатоканального пристрою Stor1

ENTER Stor2,FN\$Vol2_1; зайняття Vol2_1 ресурсів багатоканального пристрою Stor2

ENTER Stor3,FN\$Vol3_1; зайняття Vol3_1 ресурсів багатоканального пристрою Stor3

ENTER Stor4,FN\$Vol4_1; зайняття Vol4_1 ресурсів багатоканального пристрою Stor4

* DEPART wait1; звільнення черги

RELEASE Channel1; звільнення приладу Channel1

ADVANCE FN\$Pr_time1; затримка на час обслуговування рівний значенню функції Pr_time1

LEAVE Stor1,FN\$Vol1_1; звільнення Vol1_1 ресурсів пристрою Stor1

LEAVE Stor2,FN\$Vol2_1; звільнення Vol2_1 ресурсів пристрою Stor2

LEAVE Stor3,FN\$Vol3_1; звільнення Vol3_1 ресурсів пристрою Stor3

LEAVE Stor4,FN\$Vol4_1; звільнення Vol4_1 ресурсів пристрою Stor4

*****Блок

2*****

*met2 QUEUE wait2; зайняття черги

met2 SEIZE Channel2; зайняття приладу Channel2

met_Test2 TEST E BV\$CheckS2,1; якщо значення булевої змінної 1, то йдемо далі, інакше - очікування

TEST GE FN\$MaxDel,M1,discard2; якщо час перебування в системі не перевищує допустимого значення, то йдемо далі, інакше - на мітку discard2

ENTER Stor1,FN\$Vol1_2; зайняття Vol1_2 ресурсів багатоканального пристрою Stor1

ENTER Stor2,FN\$Vol2_2; зайняття Vol2_2 ресурсів багатоканального пристрою Stor2

ENTER Stor3,FN\$Vol3_2; зайняття Vol3_2 ресурсів багатоканального пристрою Stor3

ENTER Stor4,FN\$Vol4_2; зайняття Vol4_2 ресурсів багатоканального пристрою Stor4

* DEPART wait2; звільнення черги

RELEASE Channel2; звільнення приладу Channel2

ADVANCE FN\$Pr_time2; затримка на час обслуговування рівний значенню функції Pr_time2

LEAVE Stor1,FN\$Vol1_2; звільнення Vol1_2 ресурсів пристрою Stor1

LEAVE Stor2,FN\$Vol2_2; звільнення Vol2_2 ресурсів пристрою Stor2

LEAVE Stor3,FN\$Vol3_2; звільнення Vol3_2 ресурсів пристрою Stor3

LEAVE Stor4,FN\$Vol4_2; звільнення Vol4_2 ресурсів пристрою Stor4

*****Блок

3*****

*met3 QUEUE wait3; зайняття черги

met3 SEIZE Channel3; зайняття приладу Channel3

met_Test3 TEST E BV\$CheckS3,1; якщо значення булевої змінної 1, то йдемо далі, інакше - очікування

TEST GE FN\$MaxDel,M1,discard3; якщо час перебування в системі не перевищує допустимого значення, то йдемо далі, інакше - на мітку discard3

ENTER Stor1,FN\$Vol1_3; зайняття Vol1_3 ресурсів багатоканального пристрою Stor1

ENTER Stor2,FN\$Vol2_3; зайняття Vol2_3 ресурсів багатоканального пристрою Stor2

ENTER Stor3,FN\$Vol3_3; зайняття Vol3_3 ресурсів багатоканального пристрою Stor3

ENTER Stor4,FN\$Vol4_3; зайняття Vol4_3 ресурсів багатоканального пристрою Stor4

* DEPART wait3; звільнення черги

RELEASE Channel3; звільнення приладу Channel3

ADVANCE FN\$Pr_time3; затримка на час обслуговування рівний значенню функції Pr_time3

LEAVE Stor1,FN\$Vol1_3; звільнення Vol1_3 ресурсів пристрою Stor1

LEAVE Stor2,FN\$Vol2_3; звільнення Vol2_3 ресурсів пристрою Stor2

LEAVE Stor3,FN\$Vol3_3; звільнення Vol3_3 ресурсів пристрою Stor3

LEAVE Stor4,FN\$Vol4_3; звільнення Vol4_3 ресурсів пристрою Stor4

*****Блок

4*****

*met4 QUEUE wait4; зайняття черги

met4 SEIZE Channel4; зайняття приладу Channel4

met_Test4 TEST E BV\$CheckS4,1; якщо значення булевої змінної 1, то йдемо далі, інакше - очікування

TEST GE FN\$MaxDel,M1,discard4; якщо час перебування в системі не перевищує допустимого значення, то йдемо далі, інакше - на мітку discard4

ENTER Stor1,FN\$Vol1_4; зайняття Vol1_4 ресурсів багатоканального пристрою Stor1

ENTER Stor2,FN\$Vol2_4; зайняття Vol2_4 ресурсів багатоканального пристрою Stor2

ENTER Stor3,FN\$Vol3_4; зайняття Vol3_4 ресурсів багатоканального пристрою Stor3

ENTER Stor4,FN\$Vol4_4; зайняття Vol4_4 ресурсів багатоканального пристрою Stor4

* DEPART wait4; звільнення черги

RELEASE Channel4; звільнення приладу Channel4

ADVANCE FN\$Pr_time4; затримка на час обслуговування рівний значенню функції Pr_time4

LEAVE Stor1,FN\$Vol1_4; звільнення Vol1_4 ресурсів пристрою Stor1

LEAVE Stor2,FN\$Vol2_4; звільнення Vol2_4 ресурсів пристрою Stor2

LEAVE Stor3,FN\$Vol3_4; звільнення Vol3_4 ресурсів пристрою Stor3

LEAVE Stor4,FN\$Vol4_4; звільнення Vol4_4 ресурсів пристрою Stor4

*****Блок

5*****

*met5 QUEUE wait5; зайняття черги

met5 SEIZE Channel5; зайняття приладу Channel5

met_Test5 TEST E BV\$CheckS5,1; якщо значення булевої змінної 1, то йдемо далі,
інакше - очікування

TEST GE FN\$MaxDel,M1,discard5; якщо час перебування в системі не перевищує
допустимого значення, то йдемо далі, інакше - на мітку discard5

ENTER Stor1,FN\$Vol1_5; зайняття Vol1_5 ресурсів багатоканального пристрою
Stor1

ENTER Stor2,FN\$Vol2_5; зайняття Vol2_5 ресурсів багатоканального пристрою
Stor2

ENTER Stor3,FN\$Vol3_5; зайняття Vol3_5 ресурсів багатоканального пристрою
Stor3

ENTER Stor4,FN\$Vol4_5; зайняття Vol4_5 ресурсів багатоканального пристрою
Stor4

* DEPART wait5; звільнення черги

RELEASE Channel5; звільнення приладу Channel5

ADVANCE FN\$Pr_time5; затримка на час обслуговування рівний значенню функції
Pr_time5

LEAVE Stor1,FN\$Vol1_5; звільнення Vol1_5 ресурсів пристрою Stor1

LEAVE Stor2,FN\$Vol2_5; звільнення Vol2_5 ресурсів пристрою Stor2

LEAVE Stor3,FN\$Vol3_5; звільнення Vol3_5 ресурсів пристрою Stor3

LEAVE Stor4,FN\$Vol4_5; звільнення Vol4_5 ресурсів пристрою Stor4

TRANSFER ,met_tab; перехід на met_tab

***Сегмент відкидання заявок

*discard1 DEPART wait1; звільнення черги

discard1 RELEASE Channel1; звільнення приладу Channel1

TRANSFER ,met_term; перехід на met_term

*discard2 DEPART wait2; звільнення черги

discard2 RELEASE Channel2; звільнення приладу Channel2

TRANSFER ,met_term; перехід на met_term

*discard3 DEPART wait3; звільнення черги

discard3 RELEASE Channel3; звільнення приладу Channel3

TRANSFER ,met_term; перехід на met_term

*discard4 DEPART wait4; звільнення черги

discard4 RELEASE Channel4; звільнення приладу Channel4

TRANSFER ,met_term; перехід на met_term

*discard5 DEPART wait5; звільнення черги

discard5 RELEASE Channel5; звільнення приладу Channel5

TRANSFER ,met_term; перехід на met_term

*****Другий

сервер*****

*****Блок 1 серверу

2*****

*metS1 QUEUE waitS1; зайняття черги

metS1 SEIZE ChannelS1; зайняття приладу ChannelS1

met_TS1 TEST E BV\$CheckSS1,1; якщо значення булевої змінної 1, то йдемо далі, інакше - очікування

TEST GE FN\$MaxDel,M1,discardS1; якщо час перебування в системі не перевищує допустимого значення, то йдемо далі, інакше - на мітку discardS1

ENTER StorS1,FN\$Vol1_1; зайняття Vol1_1 ресурсів багатоканального пристрою StorS1

ENTER StorS2,FN\$Vol2_1; зайняття Vol2_1 ресурсів багатоканального пристрою StorS2

ENTER StorS3,FN\$Vol3_1; зайняття Vol3_1 ресурсів багатоканального пристрою StorS3

ENTER StorS4,FN\$Vol4_1; зайняття Vol4_1 ресурсів багатоканального пристрою StorS4

* DEPART waitS1; звільнення черги

RELEASE ChannelS1; звільнення приладу ChannelS1

ADVANCE FN\$Pr_time1; затримка на час обслуговування рівний значенню функції Pr_time1

LEAVE StorS1,FN\$Vol1_1; звільнення Vol1_1 ресурсів пристрою Stor1

LEAVE StorS2,FN\$Vol2_1; звільнення Vol2_1 ресурсів пристрою Stor2

LEAVE StorS3,FN\$Vol3_1; звільнення Vol3_1 ресурсів пристрою Stor3

LEAVE StorS4,FN\$Vol4_1; звільнення Vol4_1 ресурсів пристрою Stor4

*****Блок 2 серверу

2*****

*metS2 QUEUE waitS2; зайняття черги

metS2 SEIZE ChannelS2; зайняття приладу ChannelS2

met_TS2 TEST E BV\$CheckSS2,1; якщо значення булевої змінної 1, то йдемо далі,
інакше - очікування

TEST GE FN\$MaxDel,M1,discardS2; якщо час перебування в системі не перевищує
допустимого значення, то йдемо далі, інакше - на мітку discardS2

ENTER StorS1,FN\$Vol1_2; зайняття Vol1_2 ресурсів багатоканального пристрою
StorS1

ENTER StorS2,FN\$Vol2_2; зайняття Vol2_2 ресурсів багатоканального пристрою
StorS2

ENTER StorS3,FN\$Vol3_2; зайняття Vol3_2 ресурсів багатоканального пристрою
StorS3

ENTER StorS4,FN\$Vol4_2; зайняття Vol4_2 ресурсів багатоканального пристрою
StorS4

* DEPART waitS2; звільнення черги

RELEASE ChannelS2; звільнення приладу ChannelS2

ADVANCE FN\$Pr_time2; затримка на час обслуговування рівний значенню функції
Pr_time2

LEAVE StorS1,FN\$Vol1_2; звільнення Vol1_2 ресурсів пристрою StorS1

LEAVE StorS2,FN\$Vol2_2; звільнення Vol2_2 ресурсів пристрою StorS2

LEAVE StorS3,FN\$Vol3_2; звільнення Vol3_2 ресурсів пристрою StorS3

LEAVE StorS4,FN\$Vol4_2; звільнення Vol4_2 ресурсів пристрою StorS4

*****Блок 3 серверу

2*****

*metS3 QUEUE waitS3; зайняття черги

metS3 SEIZE ChannelS3; зайняття приладу ChannelS3

met_TS3 TEST E BV\$CheckSS3,1; якщо значення булевої змінної 1, то йдемо далі, інакше - очікування

TEST GE FN\$MaxDel,M1,discardS3; якщо час перебування в системі не перевищує допустимого значення, то йдемо далі, інакше - на мітку discardS3

ENTER StorS1,FN\$Vol1_3; зайняття Vol1_3 ресурсів багатоканального пристрою StorS1

ENTER StorS2,FN\$Vol2_3; зайняття Vol2_3 ресурсів багатоканального пристрою StorS2

ENTER StorS3,FN\$Vol3_3; зайняття Vol3_3 ресурсів багатоканального пристрою StorS3

ENTER StorS4,FN\$Vol4_3; зайняття Vol4_3 ресурсів багатоканального пристрою StorS4

* DEPART waitS3; звільнення черги

RELEASE ChannelS3; звільнення приладу ChannelS3

ADVANCE FN\$Pr_time3; затримка на час обслуговування рівний значенню функції Pr_time3

LEAVE StorS1,FN\$Vol1_3; звільнення Vol1_3 ресурсів пристрою StorS1

LEAVE StorS2,FN\$Vol2_3; звільнення Vol2_3 ресурсів пристрою StorS2

LEAVE StorS3,FN\$Vol3_3; звільнення Vol3_3 ресурсів пристрою StorS3

LEAVE StorS4,FN\$Vol4_3; звільнення Vol4_3 ресурсів пристрою StorS4

*****Блок 4 серверу

2*****

*metS4 QUEUE waitS4; зайняття черги

metS4 SEIZE ChannelS4; зайняття приладу ChannelS4

met_TS4 TEST E BV\$CheckSS4,1; якщо значення булевої змінної 1, то йдемо далі, інакше - очікування

TEST GE FN\$MaxDel,M1,discardS4; якщо час перебування в системі не перевищує допустимого значення, то йдемо далі, інакше - на мітку discardS4

ENTER StorS1,FN\$Vol1_4; зайняття Vol1_4 ресурсів багатоканального пристрою StorS1

ENTER StorS2,FN\$Vol2_4; зайняття Vol2_4 ресурсів багатоканального пристрою StorS2

ENTER StorS3,FN\$Vol3_4; зайняття Vol3_4 ресурсів багатоканального пристрою StorS3

ENTER StorS4,FN\$Vol4_4; зайняття Vol4_4 ресурсів багатоканального пристрою StorS4

* DEPART waitS4; звільнення черги

RELEASE ChannelS4; звільнення приладу ChannelS4

ADVANCE FN\$Pr_time4; затримка на час обслуговування рівний значенню функції Pr_time4

LEAVE StorS1,FN\$Vol1_4; звільнення Vol1_4 ресурсів пристрою StorS1

LEAVE StorS2,FN\$Vol2_4; звільнення Vol2_4 ресурсів пристрою StorS2

LEAVE StorS3,FN\$Vol3_4; звільнення Vol3_4 ресурсів пристрою StorS3

LEAVE StorS4,FN\$Vol4_4; звільнення Vol4_4 ресурсів пристрою StorS4

*****Блок 5 серверу

2*****

*metS5 QUEUE waitS5; зайняття черги

metS5 SEIZE ChannelS5; зайняття приладу ChannelS5

*met_TS5 TEST E BV\$CheckSS5,1; якщо значення булевої змінної 1, то йдемо далі, інакше - очікування

TEST GE FN\$MaxDel,M1,discardS5; якщо час перебування в системі не перевищує допустимого значення, то йдемо далі, інакше - на мітку discardS5

ENTER StorS1,FN\$Vol1_5; зайняття Vol1_5 ресурсів багатоканального пристрою StorS1

ENTER StorS2,FN\$Vol2_5; зайняття Vol2_5 ресурсів багатоканального пристрою StorS2

ENTER StorS3,FN\$Vol3_5; зайняття Vol3_5 ресурсів багатоканального пристрою StorS3

ENTER StorS4,FN\$Vol4_5; зайняття Vol4_5 ресурсів багатоканального пристрою StorS4

* DEPART waitS5; звільнення черги

RELEASE ChannelS5; звільнення приладу ChannelS5

ADVANCE FN\$Pr_time5; затримка на час обслуговування рівний значенню функції Pr_time5

LEAVE StorS1,FN\$Vol1_5; звільнення Vol1_5 ресурсів пристрою StorS1

LEAVE StorS2,FN\$Vol2_5; звільнення Vol2_5 ресурсів пристрою StorS2

LEAVE StorS3,FN\$Vol3_5; звільнення Vol3_5 ресурсів пристрою StorS3

LEAVE StorS4,FN\$Vol4_5; звільнення Vol4_5 ресурсів пристрою StorS4

TRANSFER ,met_tab; перехід на met_tab

***Сегмент відкидання заявок

*discardS1 DEPART waitS1; звільнення черги

discardS1 RELEASE ChannelS1; звільнення приладу ChannelS1

TRANSFER ,met_term; перехід на met_term

*discardS2 DEPART waitS2; звільнення черги

discardS2 RELEASE ChannelS2; звільнення приладу ChannelS2

TRANSFER ,met_term; перехід на met_term

*discardS3 DEPART waitS3; звільнення черги

discardS3 RELEASE ChannelS3; звільнення приладу ChannelS3

TRANSFER ,met_term; перехід на met_term

*discardS4 DEPART waitS4; звільнення черги

discardS4 RELEASE ChannelS4; звільнення приладу ChannelS4

TRANSFER ,met_term; перехід на met_term

*discardS5 DEPART waitS5; звільнення черги

discardS5 RELEASE ChannelS5; звільнення приладу ChannelS5

TRANSFER ,met_term; перехід на met_term

met_tab TABULATE Ttime; збір статистики

met_term TERMINATE

*****Сегмент задання часу моделювання*****

GENERATE 604800; часова одиниця - секунда, час моделювання - 604800 секунд
(7 днів)

TERMINATE 1

START 1

Содержимое dataParam.txt:

Lyambda1 FUNCTION ((AC1\3600)@24),D24; функція мат.очікування експоненційного закону розподілу інтервалу між надходженням заявок по годинно
0,1800/1,1800/2,1800/3,1800/4,1800/5,1600/6,900/7,400/8,200/9,30/10,20/11,10/12,18/
13,20/14,15/15,30/16,20/17,35/18,26/19,10/20,16/21,17/22,25/23,45

LyambD1 FUNCTION (AC1\86400),D7; функція коефіцієнту експоненційного закону розподілу інтервалу між надходженням заявок по дням

0,1/1,2/2,2/3,2/4,1/5,2/6,2

Lyambda2 FUNCTION ((AC1\3600)@24),D24; функція мат.очікування експоненційного закону розподілу інтервалу між надходженням заявок по годинно
0,1800/1,1800/2,1800/3,1800/4,1800/5,1600/6,900/7,400/8,200/9,30/10,20/11,10/12,18/
13,20/14,15/15,30/16,20/17,35/18,26/19,10/20,16/21,17/22,25/23,45

LyambD2 FUNCTION (AC1\86400),D7; функція коефіцієнту експоненційного закону розподілу інтервалу між надходженням заявок по дням

0,1/1,2/2,2/3,2/4,1/5,2/6,2

Lyambda3 FUNCTION ((AC1\3600)@24),D24; функція мат.очікування експоненційного закону розподілу інтервалу між надходженням заявок по годинно
0,1800/1,1800/2,1800/3,1800/4,1800/5,1600/6,900/7,400/8,200/9,30/10,20/11,10/12,18/
13,20/14,15/15,30/16,20/17,35/18,26/19,10/20,16/21,17/22,25/23,45

LyambD3 FUNCTION (AC1\86400),D7; функція коефіцієнту експоненційного закону розподілу інтервалу між надходженням заявок по дням

0,1/1,2/2,2/3,2/4,1/5,2/6,2

Lyambda4 FUNCTION ((AC1\3600)@24),D24; функція мат.очікування експоненційного закону розподілу інтервалу між надходженням заявок по годинно
0,1800/1,1800/2,1800/3,1800/4,1800/5,1600/6,900/7,400/8,200/9,30/10,20/11,10/12,18/
13,20/14,15/15,30/16,20/17,35/18,26/19,10/20,16/21,17/22,25/23,45

LyambD4 FUNCTION (AC1\86400),D7; функція коефіцієнту експоненційного закону розподілу інтервалу між надходженням заявок по дням

0,1/1,2/2,2/3,2/4,1/5,2/6,2

OnOff FUNCTION ((AC1\3600)@24),D24; функція доступності другого серверу погодинно

0,0/1,0/2,0/3,0/4,0/5,0/6,0/7,0/8,0/9,1/10,1/11,1/12,1/13,1/14,1/15,1/16,1/17,0/18,0/19,1/20,1/21,1/22,0/23,0

ДОДАТОК 2

"ЗАТВЕРДЖУЮ"

Начальник департаменту стратегії та розвитку нових напрямків бізнесу
ТОВ «лайфселл»

11.09.18 Лебедєв А.В.
"17" вересня 2018 р.

АКТ

впровадження результатів дисертаційної роботи
Скулиш Марії Анатоліївни

Ми, що нижче підписалися, представники Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського»:

- завідувач кафедри ІТМ д.т.н., проф. Глоба Л.С.,
- доцент кафедри ІТМ к.т.н., с.н.с. Скулиш М.А.

та представники ТОВ «лайфселл»:

- начальник департаменту стратегії та розвитку нових напрямків бізнесу Лебедєв А.В.,
- менеджер проектів з розвитку нових напрямків бізнесу Комардіна О.О.

склали цей акт, який засвідчує, що результати дисертаційної роботи Скулиш Марії Анатоліївни, а саме архітектурні рішення, протоколи взаємодії та алгоритми технології керування гібридними сервісами у телекомунікаційних системах нового покоління, апробовано в лабораторіях ТОВ «лайфселл» під час створення гнучких добре масштабованих мереж оператора мобільного зв'язку із повним або частковим застосуванням принципу віртуалізації мережевих функцій, що дозволило зменшити коефіцієнт простою обчислювальних ресурсів на 15%, уникнути зниження якості обслуговування під час сплесків перевантаження, утримати показники QoS на заданому рівні.

ТОВ «лайфселл»
Менеджер проектів з розвитку нових напрямків бізнесу

Комардіна О.О.
(підпис)
17.09.18
(дата)

КПІ імені Ігоря Сікорського
Зав. кафедри ІТМ, д.т.н., проф.

Глоба Л.С.
Докторант кафедри ІТМ, к.т.н., с.н.с.

Скулиш М.А.
17.09.18
(дата)



УКРАЇНА
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНА УСТАНОВА
НАЦІОНАЛЬНИЙ АНТАРКТИЧНИЙ НАУКОВИЙ ЦЕНТР

01601, м. Київ, бульв. Тараса Шевченка, 16, тел/факс: (380 44) 246-38-80, E-mail: uac@uac.gov.ua www.uac.gov.ua

Вір 10.12.18р. № 3
На № _____

ЗАТВЕРДЖУЮ

Т.в.о. директора Державної установи

Національний антарктичний науковий центр

І.В.МОРОЗ

2018 р.



АКТ

впровадження результатів науково-дослідної роботи на тему «Систематизація даних антарктичних досліджень з використанням інтелектуальних засобів представлення знань»

Ми, що нижче підписалися, представники Державної установи Національний антарктичний науковий центр к.б.н. Маланчук В.М., Леонов М.А. та Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» д.т.н., професор Глоба Л.С., к.т.н., н.с. Новогрудська Р.Л., к.т.н., н.с. Скулиш М.А., м.н.с. Суліма С.В. склали цей акт, який засвідчує, що було відпрацьовано програмне забезпечення щодо структуризації та систематизації даних та знань антарктичних досліджень, протоколи підтримки єдиного гетерогенного середовища, яке складається з віртуальних та фізичних ресурсів підтримки функціонування Національного центру Антарктичних даних та вдосконалено програмне забезпечення порталу «Національний центр Антарктичних даних».

Використання моделей та методів систематизації даних антарктичних досліджень дозволило організувати представлення антарктичних даних і знань, а також їх взаємодію в рамках єдиного інформаційно-комунікаційного середовища з передачею результатів антарктичних досліджень на портал NASA. Застосування методу пошуку антарктичних даних, який базується на запропонованій онтологічній моделі та враховує семантичну зв'язність елементів єдиного інформаційно-комунікаційного середовища результатів антарктичних досліджень, дозволив задавати пошуковий запит та інструменти для фільтрації знань у відповідності до метаданих, які характеризують потрібні аспекти наукового дослідження.

Розроблено механізм керування обслуговуванням гібридних сервісів у гетерогенному середовищі для обробки антарктичних даних, а також відповідне програмне забезпечення, яке дозволяє вирішити такі задачі:


- забезпечити постійний контроль за показниками якості обслуговування запитів (затримка у обслуговуванні і ймовірність успішного обслуговування запитів);
- організувати динамічний розподіл обчислювальних ресурсів телекомунікаційної системи, що дозволяє утримувати показники коефіцієнту використання ресурсів у заданих межах.

Під час впровадження роботи використано наукові результати, отримані в дисертаційних роботах к.т.н., н.с. Скулиш М.А., м.н.с. Суліми С.В.

Було відпрацьовано програмно-технічні засоби структуризації та систематизації даних та знань на платформі надання послуг інформування наукової спільноти щодо досліджень, які проводяться в Антарктиці, що дозволило підвищити ефективність роботи із накопиченням, збереженням та обробкою антарктичних даних станції «Академік Вернадський».

Від ДУ НАНЦ

 Маланчук В.М.

 Леонов М.А.

Від КНІ ім. Ігоря Сікорського

 Глоба Л.С.

 Новогрудська Р.Л.

 Скулиш М.А.

 Суліма С.В.



ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ «ХУАВЕЙ УКРАЇНА»
LIMITED LIABILITY COMPANY "HUAWEI UKRAINE"

03124Київ, вул. Радищева, буд. 10/14, (ідентифікаційний код: 33501267), тел. 498-77-18, факс 498-77-19
10/14, Radischeva street, Kiev, 03124Ukraine, (Company Code: 33501267), tel. 498-77-18, fax 498-77-19

м. Київ/ Kyiv

Вихідний № 28 від 15.01.2019 р.

"ЗАТВЕРДЖУЮ"

Керівник напрямку Safe & Smart City,

ООО «Хуавей -Україна»

О.М. Дяденко

15.01.18 р.



АКТ

впровадження результатів дисертаційної роботи
Скулиш Марії Анатоліївни

Ми, що нижче підписалися, представники Національного технічного університету України
«Київський політехнічний інститут ім. І. Сікорського»:

- завідувач кафедри ІТМ д.т.н., проф. Глоба Л.С.,
- докторант кафедри ІТМ к.т.н., с.н.с. Скулиш М.А.

та представник компанії ООО «Хуавей -Україна»:

- керівник напрямку Safe & Smart City, ООО «Хуавей -Україна» О.М. Дяденко,
- менеджер з продукту Safe & Smart City, ООО «Хуавей -Україна»

склали цей акт, який засвідчує, що результати дисертаційної роботи Скулиш Марії Анатоліївни, а саме запропоновану архітектуру телекомунікаційної мережі з віртуалізацією мережевих функцій із забезпеченням контролю якості обслуговування гібридних телекомунікаційних сервісів та керуванням обчислювальними ресурсами телекомунікаційної мережі апробовано під час побудови сучасних систем типу SmartCity, M2M, D2D в лабораторіях компанії "Huawei Ukraine", що дозволило ефективно перерозподіляти навантаження між групою фізичних та віртуальних вузлів обслуговування та покращити коефіцієнт використання обчислювальних ресурсів, а саме зменшити відсоток запитів, які було втрачено через перевищення допустимого часу обслуговування на 5%, зменшити відсотки часу низької завантаженості обчислювальних ресурсів на 8%, а відповідно їх перевантаження на 10%.

ООО «Хуавей -Україна»
Керівник напрямку Safe & Smart City,

к.т.н.

О.М. Дяденко

(підпис)

15.01.18

(дата)



КПІ імені Ігоря Сікорського
Зав. кафедри ІТМ, д.т.н., проф.

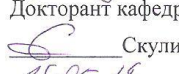
Глоба Л.С.

Докторант кафедри ІТМ, к.т.н, с.н.с.

Скулиш М.А.

15.01.18

(дата)





ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ «ХУАВЕЙ УКРАЇНА»
LIMITED LIABILITY COMPANY "HUAWEI UKRAINE"

03124, Київ, вул. Радищева, буд. 10/14, (ідентифікаційний код: 33501267), тел. 498-77-18, факс 498-77-19
10/14, Radischeva street, Kiev, 03124, Ukraine, (Company Code: 33501267), tel. 498-77-18, факс 498-77-19

"ЗАТВЕРДЖУЮ"



Керівник напрямку Safe & Smart City
ООО «Хуавей Україна»
О.М. Дяденко
"09" квітня 2018 р.

АКТ

впровадження результатів дисертаційної роботи Скулиш Марії Анатоліївни на
тему «**Методологія керування процесом обслуговування у гетерогенному
телекомунікаційному середовищі**»

Ми, що нижче підписалися, представники Національного технічного університету
України «Київський політехнічний інститут ім. І. Сікорського»:

- завідувач кафедри ІТМ д.т.н., проф. Глоба Л.С.,
- докторант кафедри ІТМ к.т.н., с.н.с. Скулиш М.А.

та представники компанії ООО «Хуавей Україна»:

- керівник напрямку Safe & Smart City ООО «Хуавей Україна» О.М. Дяденко
- менеджер з продукту Safe & Smart City ООО «Хуавей Україна» Н.А Гвоздецька

склали цей акт, який засвідчує, що результати дисертаційної роботи Скулиш Марії
Анатоліївни на тему «Методологія керування процесом обслуговування у гетерогенному
телекомунікаційному середовищі», а саме методологія забезпечення якості обслуговування
сервісів у гетерогенному телекомунікаційному середовищі, була використана при розробці
проектних рішень та засобів під час проектування систем Smart City на підприємстві
ООО «Хуавей Україна».

ООО «Хуавей Україна»
Керівник напрямку Safe & Smart City, к.т.н.
О.М. Дяденко
(підпис)

Менеджер з продукту Safe & Smart City
Н.А Гвоздецька
(підпис)
09.04.2018
(дата)

КПІ імені Ігоря Сікорського
Зав. кафедри ІТМ, д.т.н., проф.
Глоба Л.С.
(підпис)
Докторант кафедри ІТМ, к.т.н., с.н.с.
Скулиш М.А.
(підпис)
09.04.2018
(дата)